

**Вероятность,
математическая статистика,
случайные процессы**

Учебное пособие

Д. Х. Муштари

Казанский университет
механико-математический факультет

11.05.2011, версия
которую следует дополнить,
кое-где переписать и проверить

Содержание

ВВЕДЕНИЕ	3
Часть 1. СОБЫТИЯ	5
§1. Классическая модель теории вероятностей	4
§2. Геометрическая модель теории вероятностей	7
§3. Вероятностные формулы. Условная вероятность	9
§4. Независимость	13
Часть 2. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ	22
§5. Случайные величины	22
§6. Случайные векторы, наборы случайных величин	26
Решение задач. 1.	37
§7. Характеристики случайных величин	43
§8. Задача регрессии	51
Часть 3. ПОСЛЕДОВАТЕЛЬНОСТИ СЛУЧАЙНЫХ ВЕЛИЧИН	69
§9. Сходимость случайных величин	54
§10. Применение закона больших чисел — метод Монте-Карло	59
§11. Усиленный закон больших чисел	62
§12. Слабая сходимость распределений	69
§13. Характеристические функции	75
§14. Теорема Линдеберга	83
§15. Применения предельных теорем. Многомерные предельные теоремы	87
Часть 4. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА	90
§16. Выборка	90
§17. Вероятностная и статистическая модели	91
§18. Оценка параметров	94
§19. Доверительные интервалы	106
§20. Достаточные статистики	111
§21. Сравнение двух гипотез	112
§22. Задача проверки гипотез	113
§23. Обзор статистических критериев	119
§24. Проверка независимости	120
§25. Различение двух гипотез методом последовательного анализа Вальда	124
§26. Равномерно наиболее мощные критерии	124
§27. Многомерный анализ (обзор)	142
Часть 5. СЛУЧАЙНЫЕ ПРОЦЕССЫ	127
§28. Случайное блуждание	127
§29. Цепи Маркова	134
§30. Два замечательных процесса с непрерывным временем	138
§31. Процессы массового обслуживания	138
§32. Свойства траекторий винеровского процесса	145
§33. Диффузионные процессы	145
§34. Стохастические интегралы и дифференциальные уравнения	148
§35. Мартингалы	161
§36. Оптимальный момент остановки. Задача о разборчивой невесте	165
§37. Стационарные случайные процессы. Прогноз	168
ДОПОЛНЕНИЕ	168
НЕОБХОДИМЫЕ СВЕДЕНИЯ ИЗ ДРУГИХ МАТЕМАТИЧЕСКИХ КУРСОВ	169
РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА	169
ЭКЗАМЕНЫ И ЗАЧЕТЫ	179
Экзамен по теории вероятностей	179
Экзамен по математической статистике	180
Зачет по теории случайных процессов	181
ТАБЛИЦЫ ВЕРОЯТНОСТЕЙ	200

Читатель должен быть постоянно вооружен терпением, бумагой и анализом.

Кроме того, в математике, как в любом другом виде поэзии, читатель
должен быть поэтом в душе.

Мишель Лоэв. *Предисловие к книге 'Теория вероятностей'*

ВВЕДЕНИЕ

Теория вероятностей, это часть математики, которая предоставляет математический аппарат для исследования случайных явлений.

Ниже я привожу определение из авторитетного источника, но заучивать определения не надо.

Теория вероятностей — математическая наука, изучающая математические модели случайных явлений. Теория вероятностей позволяет по вероятностям одних случайных событий находить вероятности других случайных событий, связанных каким-либо образом с первыми. Это изучение основано на том, что массовые случайные явления в стационарных условиях обладают закономерностью, называемой статистической устойчивостью частот.

Из статьи *Вероятностей теория. Ю.В. Прохоров, Б.А. Севастьянов* в энциклопедии 'Вероятность и математическая статистика' (Изд-во БРЭ, 1999).

Рекомендация. Необходимо во время разбираться во всех вводимых определениях. Знаком (?) или (?n) я рекомендую постараться проверить сформулированное в тексте утверждение.

1. Частотная интерпретация вероятности события.

Пусть проводится последовательность одинаковых экспериментов в одинаковых условиях. В результате каждого эксперимента либо происходит либо не происходит некоторое событие A. Обозначим через n число экспериментов, проведенных к данному моменту времени, а через n_A — число тех экспериментов, в которые произошло событие A. Вероятностью $\mathbb{P}A$ события A называется предел частоты выпадения события A, то есть

$$\mathbb{P}A = \lim_n \frac{n_A}{n}.$$

Это определение имеет целый ряд недостатков как с точки зрения математики, так и с точки зрения практики. Определение не может считаться математическим, так как не определены использованные в нем термины — 'событие', 'эксперимент', 'одинаковые условия' и т. д. С практической точки зрения определение слишком узкое: условия проведения всех случайных экспериментов не могут оставаться постоянными (например, эксперименты с погодой, которые ставит природа, меняются в результате потепления климата), кроме того, мы можем лишь мыслить в терминах бесконечной последовательности экспериментов.

Итак, наше определение является лишь интерпретацией вероятности. Однако оно дает некую информацию о свойствах вероятности:

- 1) $n_A \geq 0$, поэтому должно быть и $\mathbb{P}A \geq 0$,
- 2) $n_A \leq n$, поэтому $\mathbb{P}A \leq 1$,
- 3) если события A и B не могут происходить одновременно, то $n_{A+B} = n_A + n_B$, где событие A + B означает, что i) произошло или A или B, ii) A и B несовместны (не могут происходить одновременно). Поэтому должно выполняться равенство

$$\mathbb{P}(A + B) = \mathbb{P}A + \mathbb{P}B.$$

В курсе мы будем иметь дело с аксиомами вероятности (см. ниже), а также с различными моделями, в рамках которых мы сможем иногда вычислять вероятности. Тем не менее, бывает полезно представить себе то, что описывается этими моделями. Приведу простой пример — анекдот, возможно из преподавательской практики, а может и нет, который мне рассказывали

мои коллеги, преподающие математику гуманитариям и даже студентам технических специальностей. Студента спрашивают — сколько будет, если разделить 1/2 на 3? Ответ: 6. Но если спросить — сколько получится, если поллитра водки разделить на троих, ответ будет более разумным — немного меньше одного стакана. Таким образом, мыслить в рамках реальной ситуации бывает полезно. В курсе рекомендуется мыслить в рамках частотной интерпретации вероятности при анализе Примера 1 в разделе 'Частотная интерпретация' и при решении других задач.

2. Аксиоматика теории вероятностей

Математическую аксиоматику для теории вероятностей предложил А.Н. Колмогоров (используя развитую А. Лебегом и другими математиками теорию меры).

Вероятностное пространство — это тройка $(\Omega, \mathcal{A}, \mathbb{P})$, где Ω называется *пространством элементарных исходов*, точки Ω мы будем называть *элементарными исходами* и обозначать ω , \mathcal{A} — σ -алгебра событий — подмножество в Ω , \mathbb{P} — *вероятность* на \mathcal{A} , т.е.

- i) $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$,
- ii) $\mathbb{P}\left(\sum_n A_n\right) = \sum_n \mathbb{P}(A_n)$ для любой последовательности попарно несовместных событий A_n ,
- iii) $\mathbb{P}(\Omega) = 1$.

Итак, под 'событием' мы понимаем множество из σ -алгебры \mathcal{A} , 'эксперимент' — это выпадение некоторого $\omega \in \Omega$, 'событие A произошло' означает, что $\omega \in A$. Событие $A \cup B$ означает ' A или B ' (это событие произошло, если произошло событие A или событие B), при этом то же самое обозначается $A + B$, если события A и B несовместны; AB означает ' A и B ' (в теории множеств или теории меры используется более новое обозначение $A \cap B$, знак произведения более старый, но в теории вероятности оказывается очень удобным, впрочем иногда мы будем дублировать знак \cap , чтобы подчеркнуть наличие пересечения множеств), $AB = \emptyset$ означает ' A и B несовместны' (одновременно происходить не могут, т.е. ω не может попасть одновременно в оба этих события); $A \setminus B$ означает событие ' A , но не B '; \emptyset — *невозможное событие*; Ω — *достоверное событие*. Дополнение события A обозначается \bar{A} и означает событие 'не A '. События мы часто будем обозначать фигурными скобками — {...}, а внутри скобок будут записываться определяющие событие условия. Для нескольких событий используются также обозначения $\sum_i, \bigcup_i, \prod_i$. Свойство $A \subset B$ звучит как 'событие A влечет событие B ' (имеется в виду, что если элементарный исход ω принадлежит A , то он принадлежит B).

Скобки в обозначениях $\mathbb{P}(A)$ (а также в будущем в обозначениях $\mathbb{E}(\xi)$ и $\mathbb{D}(\xi)$) мы часто будем опускать.

Заметим, что по сравнению с проведенным выше анализом частотной интерпретации вероятности мы усилили свойство *аддитивности*, заменили его на σ -аддитивность.

В рамках развитой на основе этой аксиоматики теории будут введены те термины, которые были в частотной интерпретации вероятности, и будет доказана сходимость частоты выпадения события к вероятности события. Но это произойдет еще очень не скоро.

Отметим важные, но очевидные следствия аксиом. В тексте они неоднократно используются даже без объяснения. Хотя они и очевидны, читателю надо их проверить.

- i) Если $A \subset B$, то $\mathbb{P}A \leq \mathbb{P}B$. Действительно,

$$\mathbb{P}B = \mathbb{P}A + \mathbb{P}(B \setminus A), \text{ где } \mathbb{P}(B \setminus A) \geq 0.$$

- ii) $\mathbb{P}A = 1 - \mathbb{P}\bar{A}$. Иногда $\mathbb{P}\bar{A}$ вычисляется легче.

- iii) $\mathbb{P}(A \cup B) \leq \mathbb{P}A + \mathbb{P}B$. Действительно,

$$\mathbb{P}(A \cup B) = \mathbb{P}A + \mathbb{P}(B \setminus A), \mathbb{P}(B \setminus A) \leq \mathbb{P}B.$$

- iv) $\mathbb{P}A = 0, \mathbb{P}B = 0 \Leftrightarrow \mathbb{P}(A \cup B) = 0$, поэтому $\mathbb{P}A = 1, \mathbb{P}B = 1 \Leftrightarrow \mathbb{P}(AB) = 1$.

- v) Чтобы использовать формулу $\mathbb{P}(A+B) = \mathbb{P}A + \mathbb{P}B = 0$, надо проверить равенство $AB = \emptyset$.

vi) Из $\emptyset + \emptyset = \emptyset$ следует $\mathbb{P}\emptyset = 0$.

Упражнения надо попытаться решить в период подготовки к следующей лекции.

Упражнение 1. а. Построить алгебру всех подмножеств множества $\{1, 2, 3\}$, не забудьте про пустое множество. Почему эта алгебра является также и σ -алгеброй.

б. Докажите, что в любой конечной алгебре множеств любое множество алгебры содержит минимальное непустое множество алгебры. Такой элемент алгебры называется атомом. Докажите, что различные атомы не пересекаются. Построить алгебру, порожденную тремя атомами. Сколько элементов в этой алгебре?

Замечание. Полезнейшим примером вероятностного пространства является $\Omega = [0, 1]$, \mathfrak{A} — σ -алгебра всех борелевских подмножеств $[0, 1]$, \mathbb{P} — мера Лебега λ .

Часть 1. СОБЫТИЯ

§1. Классическая модель теории вероятностей

Хотя эпиграф к курсу ориентирует нас на постоянное применение математического анализа, в первой части курса мы будем иметь с простыми вещами, для понимания которых достаточно знания школьной математики.

1. Вычисление вероятности в классической модели.

В классической модели (или в модели равновероятных исходов) мы в качестве Ω берем конечное множество, в качестве \mathfrak{A} — множество всех подмножеств Ω , и будем считать равными вероятности всех элементов Ω . $|A|$ будет обозначать число элементов множества A . Тогда (в этом легко убедиться)

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} \quad (pr)$$

Словами это равенство озвучивается следующим образом: вероятность события в классической модели — это отношение числа *благоприятствующих* исходов к числу всех исходов.

Замечание. Разумеется, введенная выше вероятность в классической модели удовлетворяет всем аксиомам вероятности (на всякий случай проверьте!), в частности, она аддитивна. В некоторых курсах теории вероятностей для нематематиков этот факт называется **теоремой аддитивности**.

Итак, чтобы решить задачу на классическую модель, нужно:

- 1) придумать модель, т.е. Ω ,
- 2) выделить благоприятствующие исходы, т.е. подмножество $A \subset \Omega$,
- 3) подсчитать $|A|$ и $|\Omega|$, и поделить одно на другое.

Упражнение. Проверьте сами, что вероятность в классической модели удовлетворяет условиям i) - iii).

2. Элементы комбинаторики.

Для решения задач в рамках классической модели мы введем несколько комбинаторных объектов. Напомним, что

$$A \times B := \{(a, b) : a \in A, b \in B\}, |A \times B| = |A||B|, \quad (*)$$

$$A^n := \{(a_1, \dots, a_n) : a_i \in A \forall i\}, |A^n| = |A|^n.$$

Доказательство первого равенства не нужно, по существу оно может служить определением произведения двух натуральных чисел: в каждом из множеств $\{a\} \times B$ имеется $|B|$ элементов, а $A \times B$ является суммой $|A|$ экземпляров таких множеств. Впрочем, мы все-таки объясним это подробнее. Будем считать очевидным равенства $|B_1 + B_2 + \dots + B_r| = |B_1| + |B_2| + \dots + |B_r|$. Обозначим $A = \{a_1, \dots, a_r\}$. Тогда по определению $A \times B$ имеем:

$$A \times B = \{a_1\} \times B + \dots + \{a_r\} \times B.$$

Вычисляем:

$$|A \times B| = |\{a_1\} \times B| + \dots + |\{a_r\} \times B| = r \times |\{a_1\} \times B| = r \times |B|.$$

Второе равенство выводится из первого методом математической индукции (индукция по n), для $n = 2$ мы имеем $|A^2| = |A \times A| = |A|^2$.

Замечание. $3 \times 3 = 9$ эффективнее чем $3 \times 2 = 6$. Первое равенство мы можем записать через число пар элементов из одного множества:

$$\begin{aligned} 3 \times 3 &= |\{a, b, c\}| \times |\{a, b, c\}| = \\ &= |\{a, b, c\} \times \{a, b, c\}| = |\{(a, a), (a, b), (a, c), (b, a), (b, b), (b, c), (c, a), (c, b), (c, c)\}| = 9, \end{aligned}$$

второе равенство — число пар из произведения двух разных множеств:

$$3 \times 2 = |\{a, b, c\}| \times |\{x, y\}| = \\ = |\{a, b, c\} \times \{x, y\}| = |\{(a, x), (a, y), (b, x), (b, y), (c, x), (c, y)\}| = 6.$$

Однако в качестве второго множества мы можем взять часть $\{a, b, c\}$:

$$3 \times 2 = |\{a, b, c\}| \times |\{b, c\}| = |\{a, b, c\} \times \{b, c\}| = |\{(a, b), (a, c), (b, b), (b, c), (c, b), (c, c)\}| = 6.$$

Но эта формула некрасива и несимметрична, из каких-то соображений мы выделили элемент a . Теперь перейдем к степеням — методом математической индукции доказывается, что

$$|A^n| = |A|^n,$$

действительно, используя предположение индукции, мы получаем

$$|A^n| = |A^{n-1} \times A| = |A^{n-1}| \times |A| = |A|^{n-1} \times |A| = |A|^n.$$

Напомним, что A^n — это множество всех n -к (a_1, a_2, \dots, a_n), элементы которых пробегают множество A , выше мы назвали такие n -ки упорядоченными выборками с возвращениями длины n из множества A . В анализе и алгебре вы сталкивались с n -мерным пространством \mathbf{R}^n , которое состоит из всех точек $x = (x_1, x_2, \dots, x_n)$, а число x_i называлось i -й координатой точки x . Наша ситуация отличается тем, что множество A конечно, поэтому мы можем подсчитать число элементов A^n .

Элементы A^n , где A состоит из r элементов, мы будем называть *упорядоченными выборками длины n с возвращениями из r -элементного множества*. Таким образом произвольный элемент (a_1, \dots, a_n) в A^n мы считаем задаваемым в процессе выбора n элементов из A , при этом фиксируется номер каждого выбранного элемента и после выбора этот элемент возвращается назад (и может быть выбран снова под другими номерами). Напомним, что число таких выборок равно r^n . Кроме таких выборок, мы будем также рассматривать *упорядоченные выборки длины n без возвращений из r -элементного множества*. Число таких выборок равно (если $n \leq r$)

$$r(r-1)\dots(r-n+1) = \frac{r!}{(r-n)!}.$$

Это утверждение доказывается индукцией по n — к каждой выборке длины n мы можем дописать один элемент из оставшихся $r - n$ элементов, которые еще не использовались в ней раньше, поэтому при переходе от n к $n + 1$ число таких выборок увеличивается в $r - n$ раз. Таким образом наше равенство остается верным и для $n = n + 1$ (если по прежнему $n + 1 \leq r$, в противном случае процесс выбора заканчивается и мы получаем число $r!$ всех перестановок нашего множества).

Учитывая, что в школьной программе комбинаторики больше нет, мы остановимся на этой выкладке подробнее. Замечу, что во многих книгах упорядоченные выборки без возвращений называются *размещениями n элементов в r местах*. Итак, мы обобщаем понятие произведения $A \times B$ и с каждым элементом $a \in A$ связываем множество B_a , рассматривая множество пар вида $\sum_{a \in A} \{a\} \times B_a$. Разумеется, число элементов этого множества $A^{(2)}$ равно $|A^{(2)}| = \sum_{a \in A} |B_a|$. Если же все множества B_a состоят из одинакового числа элементов, то число элементов $A^{(2)}$ равно $|A| \times |B_a|$, где $a \in A$ произвольно. Эти рассуждения применимы к подсчету числа выборок длины 2 из множества A . Такие выборки представляют собой пары (a, b) , где $a \in A$, $b \in B_a = A \setminus \{a\}$ (b не может равняться a). Если $|A| = r$, то $|B_a| = r - 1$, и общее число таких пар равно $r(r - 1)$. Теперь перейдем к случаю выборок длины 3. Такие выборки являются элементами множества $A^{(3)} = \sum_{(a,b) \in A^{(2)}} \{(a, b)\} \times B_{(a,b)}$, где $B_{(a,b)} = A \setminus \{a, b\}$, и состоит из $r - 2$ элементов.

Другими словами, из каждой выборки длины 2 можно сделать $r - 2$ выборок длины 3, общее

число таких выборок равно $r(r-1)(r-2)$. Случай произвольного $n \leq r$ обосновывается методом математической индукции.

Обозначим через $A^{(n)}$ множество всех упорядоченных выборок без возвращений длины n из множества A , где $|A| = r$. Согласно предположению индукции $|A^{(n)}| = r(r-1)...(r-n+1)$. Теперь запишем представление для множества $A^{(n+1)}$:

$$A^{(n+1)} = \sum_{\{a_1, \dots, a_n\} \in A^{(n)}} \{a_1, \dots, a_n\} \times (A \setminus \{a_1, \dots, a_n\}). \quad (*)$$

Очевидно, что $|A \setminus \{a_1, \dots, a_n\}| = r - n$ для любого $\{a_1, \dots, a_n\} \in A^{(n)}$ (элементы $A^{(n)}$ — подмножества A), поэтому

$$|A^{(n+1)}| = \underbrace{(r-n) + \dots + (r-n)}_{|A_n| \text{ times}} = |A^{(n)}| \times (r-n) = r(r-1)...(r-n+1)(r-n).$$

Замечание. Еще раз рассмотрим множество всех упорядоченных выборок без возвращений длины n из множества A , где $|A| = r$. Очевидно, что для непустоты этого множества необходимо $n \leq r$. Упорядоченные выборки без возвращений часто называются *размещениями* — мы размещаем числа $1, 2, \dots, n$ в множестве A так, что разные числа $i \neq j$ размещаются в разных элементах $a_i \neq a_j$. Мы можем такие размещения интерпретировать также как выборки — если число i размещено в элементе a_i , то мы говорим, что на i -м шаге выбрано a_i . Число всех упорядоченных выборок без возвращений длины n из множества A , где $|A| = r$, равно $r(r-1)...(r-n+1)$. Эта формула доказывается также методом математической индукции. Обозначим через $A(a_1, \dots, a_s)$ множество A , из которого мы выкинули (разные!) элементы a_1, \dots, a_s . Обозначим через $A^{(s)}$ множество всех упорядоченных выборок без возвращений длины s из множества A .

Если $n = r$, то упорядоченные выборки называются *перестановками*, их число равно $r!$. Дело в том, что в этом случае не производится выбор, так как должны быть выбраны все элементы. Поэтому наша упорядоченная выборка длины r лишь фиксирует порядок выбора элементов A , переставляя их новых образом.

Наконец, число всех *неупорядоченных выборок длины n без возвращений из r -элементного множества* (во многих учебниках они называются *сочетаниями*) равно

$$C_r^n = \frac{r!}{n!(r-n)!}. \quad (**)$$

Действительно, в каждой неупорядоченной выборке порядок можно задать $n!$ способами, то есть сделать из нее столько упорядоченных выборок без возвращений. Поэтому число неупорядоченных выборок длины n без возвращения в $n!$ раз меньше числа всех упорядоченных выборок без возвращения.

Объясним это подробнее. Равенство (*) можно записать и по другому: $|A| = \frac{|A \times B|}{|B|}$. Любую упорядоченную выборку без возвращений мы можем интерпретировать как пару (α, β) , где α — неупорядоченная выборка длины n без возвращений из r -элементного множества (порядок n выбранных элементов не фиксируется) и β — некоторый способ упорядочения n выбранных элементов. Как мы установили, число пар (α, β) равно $\frac{r!}{(r-n)!}$, а число способов упорядочения равно $n!$.

Все это можно объяснить и по другому. В множестве всех упорядоченных выборок мы вводим отношение эквивалентности: две выборки считаются эквивалентными, если они содержат одни и те же элементы множества, из которого производится выбор. Очевидно, что в каждом классе эквивалентности одинаковое число элементов, и каждый класс эквивалентности является неупорядоченной выборкой. Число классов эквивалентности равно отношению общего числа элементов, деленного на число элементов в каждом классе.

В ряде задач в качестве пространства Ω элементарных исходов мы можем использовать как множество всех упорядоченных выборок без возвращения фиксированной длины n , так и множество всех неупорядоченных выборок той же длины. Во второй модели число элементарных исходов сокращается в $n!$ раз, но во столько же раз сокращается и число благоприятствующих исходов. Таким образом, число элементов Ω уменьшается, модель становится более компактной, а нужная нам вероятность не меняется. Однако использовать ту же идею уменьшения для модели с упорядоченными выборками с возвращениями нельзя, даже если порядок элементов выборки никак не связан с интересующим нас событием. Дело в том, что в классах упорядоченных выборок, соответствующих одной неупорядоченной выборке, может оказаться разное число элементов, и при переходе от одной модели к другой вероятность по формуле (pr) изменится. Приведем пример: из одной неупорядоченной выборки (a, b, c) получается 6 упорядоченных выборок $(a, b, c), (b, c, a), (c, a, b), (a, c, b), (c, b, a), (b, a, c)$, в то же время из выборки (a, a, b) можно получить 3 упорядоченные выборки $(a, a, b), (a, b, a), (b, a, a)$, а из выборки (a, a, a) — лишь одну упорядоченную выборку.

Указанные выше множества выборок могут использоваться при решении задач как множество всех элементарных исходов или множество всех благоприятствующих исходов. При решении задачи надо проверить, что все благоприятствующие исходы — выборки того же вида, что и все исходы. Однако в дальнейшем вы увидите, что при решении некоторых задач для подсчета числа благоприятствующих исходов их выгодно интерпретировать как выборки другого вида.

Три задачи. 1. В подъезд 32-этажного дома вошли 10 человек, не связанных никакими знакомствами. Лифт обслуживает этажи с 5 по 32. Найти вероятность того, что все 10 человек воспользуются лифтом. В качестве Ω мы берем все упорядоченные выборки с возвращениями длины 10 из множества в 32 этажей. Отличие от предыдущей задачи состоит в том, что несколько человек могут выйти на одном этаже, таким образом, все элементарные исходы представляют собой выборки с возвращениями. Опять занумеруем всех пассажиров. Тогда элементы Ω оказываются упорядоченными выборками с возвращениями, число таких выборок равно 32^{10} . Точно также считаем число благоприятствующих исходов, которое состоит из упорядоченные выборки с возвращениями, но из множества в 28 этажей. Ответ: $\left(\frac{28}{32}\right)^{10}$.

Замечания. Чтобы модель была разумной, нужно, чтобы все этажи состояли из одинакового набора квартир, и не было бы ни одного ресторана. В противном случае модель будет строить трудно, брать в качестве Ω множество квартир нельзя, если квартиры имеют разную площадь. Условия отсутствия знакомств существенно при выборе модели. Если бы все 10 человек были друзьями, он шли бы на один этаж, и их всех можно было бы рассматривать как одного человека. В этом случае ответ равен $\left(\frac{28}{32}\right)$.

2. В вагоне 8 купе по 4 места. В него заходят 10 человек и рассаживаются наугад. Найти вероятность того, что в первом купе не окажется пассажиров. Здесь возможны две модели, дающие одинаковый ответ. Рассмотрим всевозможные упорядоченные выборки (мы перенумеровали пассажиров по алфавиту) без возвращения (пассажиры садятся на разные места). Число таких выборок равно $32 \cdot 31 \cdot \dots \cdot 23$. Число таких же благоприятствующих выборок равно $28 \cdot 27 \cdot \dots \cdot 19$. Итак, ответ равен

$$\frac{28 \cdot 27 \cdot \dots \cdot 19}{32 \cdot 31 \cdot \dots \cdot 23}.$$

Другое решение получается, если мы не нумеруем пассажиров. В условиях задачи эта нумерация не фигурирует. Итак, рассматриваются выборки длины 10 из множества 32 мест неупорядоченные без возвращения. Число таких выборок равно C_{32}^{10} . Благоприятствующие исходы берутся уже из множества 28 мест, их число равно C_{28}^{10} . Делим второе число на первое, получаем тот же самый ответ.

3. В лифт 8-этажного дома сели 5 человек. Найти вероятность того, что все они выйдут на разных этажах (с 2 по 8 этаж). В качестве Ω мы берем все упорядоченные выборки с возвращениями длины 5 из множества 7 этажей. Отличие от предыдущей задачи состоит в

том, что несколько человек могут выйти на одном этаже, таким образом, все элементарные исходы представляют собой выборки с возвращениями. Опять занумеруем всех пассажиров. Тогда элементы Ω оказываются упорядоченными выборками с возвращениями, число таких выборок равно 7^5 . Благоприятствующие исходы являются выборками без возвращениями, но если мы будем считать их неупорядоченными, то они не будут принадлежать Ω . Так делать мы не имеет право. Но взяв в качестве благоприятствующих исходов упорядоченные выборки без возвращения, мы получим решение задачи $\left(\frac{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3}{7^5}\right)$. Однако почему в этой задаче мы не можем использовать другую модель, состоящую из неупорядоченных выборок? Дело в том, что ответ окажется другим. Каждая неупорядоченная выборка без возвращений содержит одно и то же число упорядоченных выборок без возвращений, для выборок с возвращениями это неверно. Например, выборка $(2, 2, 2, 2, 2)$ содержит лишь одну упорядоченную выборку, а выборка $(2, 3, 4, 5, 6)$ содержит $5!$ упорядоченных выборок. Итак, если мы можем занумеровать пассажиров, почему бы это не сделать. Ω , состоящее из неупорядоченных выборок, может привести к ошибке. Ниже будет приведен пример из физики (квантовая статистика), когда нумерация в принципе невозможна, в этой ситуации правильное решение будет другим.

Замечание. Грубая ошибка — в качестве элементов Ω взять упорядоченные выборки без возвращений, а в качестве элементов A — неупорядоченные выборки без возвращений. Тогда мы не можем утверждать, что $A \subset \Omega$. А если сделать наоборот — элементы Ω брать неупорядоченными, а элементы A — упорядоченными, вероятность может оказаться даже больше $1!$

3. Пример с деталями (теоретический). Имеется N деталей, из которых R деталей стандартно, а $N - R$ деталей браковано. Выбирается n деталей. Какова вероятность того, что из них r деталей стандартно, а $n - r$ деталей браковано?

Ясно, что одну деталь мы не выбираем дважды (кстати, в условиях это можно было бы оговорить). Поэтому Ω состоит из выборок без возвращения. Никакой нумерации деталей в условиях задачи нет, поэтому мы можем использовать как модель, состоящую из упорядоченных выборок, так и модель из неупорядоченных выборок. На мой взгляд, вторая модель удобнее, так как при вычислении числа благоприятствующих исходов нам не надо будет распределить между стандартными и бракованными деталями еще и номера. Итак, $|\Omega| = C_N^n$. Каждую благоприятствующую выборку мы можем представить в виде пары $(\omega_{\text{ст}}, \omega_{\text{бр}})$, где $\omega_{\text{ст}}$ — это неупорядоченная выборка длины r из R стандартных деталей, а $\omega_{\text{бр}}$ — неупорядоченная выборка длины $n - r$ из $N - R$ бракованных деталей. Итак, наше событие A представляется как произведение двух множеств, $|A| = C_R^r C_{N-R}^{n-r}$.

4. Симметричная модель Бернулли. Симметричная монета бросается n раз, вычислить вероятность выпадения k гербов в n испытаниях.

Бросание монеты мы можем интерпретировать как выбор между гербом и решкой. Итак, мы можем использовать в качестве Ω множество всех упорядоченных выборок длины n из 2-элементного множества (герб, решка). Поэтому $|\Omega| = 2^n$. (Мы можем попытаться использовать и неупорядоченные выборки, но их использование изменит вероятность: при $n = 2$ упорядоченная модель дает нам 4 элементарных исхода — {герб, герб}, {герб, решка}, {решка, герб}, {решка, решка}, а неупорядоченная модель — только три элементарных исхода. Поэтому так делать нельзя.) Заметим, что упорядоченную модель легко обосновать: из симметрии между гербом и решкой должны следовать равенства

$$\mathbb{P}\{\text{герб, герб}\} = \mathbb{P}\{\text{герб, решка}\} = \mathbb{P}\{\text{решка, решка}\} = \mathbb{P}\{\text{решка, герб}\}.$$

В нашей модели событие — 'на i -м шаге выпал герб' мы интерпретируем как множество всех таких $n-k$ из гербов или решек, в которых на i -м месте находится герб.

Как теперь подсчитать число элементов A_k — благоприятствующих исходов. Оказывается, для подсчета эти исходы нужно переинтерпретировать. Очевидно, что каждый благоприятствующий исход задается k числами — номерами испытаний, в которых выпал герб, в остальных испытаниях выпала решка. Итак, каждый благоприятствующий исход мы интерпретируем

как выборку (n_1, \dots, n_k) длины k из множества $\{1, \dots, n\}$. Разумеется, порядок выбора чисел n_i не меняет элементарный исход в A_k , а один и тот же номер испытания не может быть выбран дважды. Итак, $|A_k| = C_n^k$,

$$\mathbb{P}A = \frac{C_n^k}{2^n}.$$

Проверка использует бином Ньютона.

$$\mathbb{P}(\Omega) = \sum_{k=0}^n \binom{n}{k} = 2^n \mathbb{P}(A_k) = \sum_{k=0}^n \binom{n}{k} = 2^n \frac{C_n^k}{2^n} = \frac{(1+1)^n}{2^n} = 1.$$

Замечание. В решении этой задачи мы отошли от ранее выдвинутого принципа — элементы множества A должны быть выборками того же вида, что и элементы множества Ω . Дело в том, что множество всех исходов Ω состоит из упорядоченных выборок с возвращениями длины n из 2-элементного множества $\{\text{герб}, \text{герб}\}$, а множество благоприятных исходов состоит из неупорядоченных выборок без возвращений, но из множества в n элементов и длины k .

Задача 4. Имеются две клетки и n кроликов, каждый из которых случайным образом размещается в одной из двух клеток. Какова вероятность того, что в первой клетке окажется ровно k кроликов? Эта задача совпадает с предыдущей и имеет то же решение, для каждого кролика мы подбрасываем монету и в зависимости от результата помещаем его в первую или вторую клетку. Но новая формулировка допускает обобщение на r клеток.

5. Статистика Больцмана — Maxwellла.

Задача с кроликами (в отличие от задачи с монетами) обобщается на случай r клеток. Итак, n кроликов наугад размещаются по r клеткам (то есть, каждый кролик наугад помещается в одну из r клеток). Какова вероятность, что в 1-й клетке окажется n_1 кролик, во 2-й — n_2 кроликов, ..., в r -й клетке — n_r кроликов? Как и в прошлой ситуации, Ω состоит из всех упорядоченных выборок длины n с возвращениями из множества в r клеток. Итак, $|\Omega| = r^n$. Для подсчета числа благоприятствующих исходов в

$$A = \{ \text{в 1-й клетке} — n_1, \text{ во 2-й} — n_2, \dots, \text{ в } r\text{-й клетке} — n_r \}$$

используется обобщение формулы для числа сочетаний,

$$|A| = \frac{n!}{n_1!n_2!\dots n_r!}.$$

Доказательство методом математической индукции по r , для перехода от $r-1$ к r мы сначала объединяем $r-1$ и r клетки, получаем событие

$$A' = \{ \text{в 1-й клетке} — n_1, \text{ во 2-й} — n_2, \dots, \text{ в } r-1\text{-й клетке} — n_{r-1} + n_r \}$$

(в новом вероятностном пространстве Ω' величины $(r-1)^n$). По предположению индукции для $r-1$ имеем

$$|A'| = \frac{n!}{n_1!n_2!\dots(n_{r-1} + n_r)!}.$$

Далее, согласно нашему утверждению для $r=2$ мы можем разделить $n_{r-1} + n_r$ кроликов между $r-1$ -й и r -й клетками в нужном соотношении ровно $\frac{(n_{r-1} + n_r)!}{n_{r-1}!n_r!}$ способами. Таким образом, каждый элемент в A' дает ровно $\frac{(n_{r-1} + n_r)!}{n_{r-1}!n_r!}$ элементов в A . Поэтому

$$|A| = \frac{n!}{n_1!n_2!\dots(n_{r-1} + n_r)!} \cdot \frac{(n_{r-1} + n_r)!}{n_{r-1}!n_r!} = \frac{n!}{n_1!n_2!\dots n_r!}.$$

Итак,

$$\mathbb{P}\{n_1, \dots, n_r\} = r^{-n} \frac{n!}{n_1!n_2!\dots n_r!}. \quad (1)$$

Полученный нами результат в физике называется статистикой Больцмана — Максвелла. Но слова нужно произносить другие. Имеется r состояний и n частиц (но не слишком маленьких). Тогда распределение частиц по состояниям описывается формулой (1).

6. Статистика Бозе — Эйнштейна.

Однако в случае совершенно идентичных квантовых частиц (предполагается, что в одном состоянии может быть несколько частиц) распределение частиц по состояниям описывается другой статистикой, которая называется статистикой Бозе — Эйнштейна.

В рамках статистики Больцмана — Максвелла обмен частицами между двумя состояниями меняет элементарный исход (частицы ведут себя как кролики).

В рамках статистики Бозе — Эйнштейна обмен частицами между двумя состояниями не меняет элементарный исход (так как частицы неразличимы между собой). В результате элементарным исходом оказывается набор чисел (n_1, \dots, n_r) , указывающих число частиц в каждом состоянии, причем все такие наборы равновероятны. Итак, в отличие от формулы (1) мы имеем

$$\mathbb{P}\{n_1, \dots, n_r\} = \frac{1}{|\Omega|}. \quad (2)$$

Для вычисления этой вероятности нам нужно научиться считать число $|\Omega|$, которое состоит из всех возможных представлений

$$n_1 + \dots + n_r = n \quad (3),$$

где n_i — целые неотрицательные числа. Идея вычисления состоит в представлении (2) в виде рисунка вида

$$|| \otimes | \otimes || \otimes \otimes | \otimes \otimes \otimes || \otimes |,$$

где каждый знак \otimes представляет частицу, а $|$ отделяет частицы в одном состоянии от частиц в другом состоянии. Например, нарисованная выше картинка означает, что $n_1 = 0$, $n_2 = 0$, $n_3 = 1$, $n_4 = 1$, $n_5 = 0$, $n_6 = 2$, $n_7 = 3$, $n_8 = 0$, $n_9 = 1$, $n_{10} = 0$. Всего в такой картинке n знаков \otimes и $r - 1$ знаков $|$. Обратно, каждая такая картинка определяет распределение n частиц по r состояниям. Число $|\Omega|$, которое совпадает с числом таких картинок, равно C_{n+r-1}^n . Вероятность каждой картинки равна $1/C_{n+r-1}^n$.

Задача 5. 2 частицы в 2 состояниях. В статистике Больцмана — Максвелла четыре элементарных исхода $12|$, $1|2$, $2|1$, $2|2$, поэтому $\mathbb{P}\{1|2 + 2|1\} = 1/2$. В статистике Бозе — Эйнштейна три элементарных исхода $\otimes \otimes |$, $\otimes | \otimes$, $| \otimes \otimes$, поэтому $\mathbb{P}\{\otimes | \otimes\} = 1/3$.

Хороший вопрос. Итак, все дело в неразличности частиц и различности монет. А что будет, если мы будем подбрасывать две монеты в кулаке. Тогда мы не сможем различать исходы {герб, решка} и {решка, герб}, не будет ли тогда $\mathbb{P}\{1 \text{ герб}, 1 \text{ решка}\} = 1/3$? Нет, так как мы можем как-то отметить две монеты, например, краской, но с одинаковыми частицами это невозможно. Точно так же ситуация в физическом эксперименте изменится, если частицы, которые мы считаем одинаковыми, на самом деле могут быть различны, состоят из двух разных сортов (даже если они очень похожи друг на друга). Тогда частота выпадения события $\{\otimes | \otimes\}$ будет лежать между числами $1/2$ и $1/3$ и будет определяться долями частиц первого и второго типа. Итак, даже не умея различать частицы первого и второго типа, мы будем приблизительно знать, что p — доля частиц одного из двух типов, $1 - p$ — доля частиц другого типа.

Замечание. Статистика Больцмана — Максвелла справедлива лишь для частиц, которые называются бозонами. Для электронов и некоторых частиц, которые называются фермионами, действует статистика Ферми — Дирака, в которой постулируется не только неразличимость частиц, но и невозможность попадания двух частиц в одно состояние. В этой статистике $r \leq n$ и $|\Omega| = C_n^r$. Вид частиц роли не играет. Чисто вероятностные соображения и наблюдения частоты не позволят здесь указать наличие среди похожих фермионов нескольких видов частиц.

Задача 6. 5 бозонов размещены в 8 состояниях. Какова вероятность того, что все они попадут в состояния с номерами 1, 2, 3, 4?

Как мы уже видели выше, Ω состоит из всех наборов вида $\cdot|\cdot|\cdot|\cdot|\cdot|\cdot|\cdot|\cdot$, где \cdot обозначает неизвестное нам некоторое число частиц (возможно, что нулевое), но общее число частиц равно 5. Множество A благоприятствующих исходов состоит из всех наборов $\cdot|\cdot|\cdot|\cdot|||$ (в последних 4 состояниях частиц заведомо нет). Как мы уже видели выше, $|\Omega| = C_{12}^5$. При подсчете числа элементов A последние 4 состояния можно не учитывать. Таким образом, A состоит из всех исходов вида $\cdot|\cdot|\cdot|\cdot$, частиц по прежнему 5. Ответ: $|A| = C_8^5$, $\mathbb{P}(A) = C_8^5/C_{12}^5$.

Упражнение 2. Имеется 4 детали, из них детали $\{a, b\}$ имеют 1 сорт и $\{x, y\}$ имеют 2 сорта. из них вытащили 2 детали. Какова вероятность, что вытащили по одной детали каждого сорта? Дайте формулу и запишите Ω — множество всех исходов и A — множество всех благоприятствующие исходов.

Упражнение 3. Имеется 12 деталей, из них 3 детали имеют 1 сорт, 4 детали имеют 2 сорт и 5 деталей имеют 3 сорт. Вытащили 6 деталей. Какова вероятность, что среди них по 2 детали каждого сорта?

§2. Геометрическая модель теории вероятностей

В геометрической модели равновероятны события, имеющие одинаковую геометрическую меру. В качестве Ω берется некоторое множество на прямой, на плоскости, в пространстве, в r -мерном пространстве, на котором задана соответственно длина, площадь, объем, r -мерный объем, а может также длина дуги на окружности, длина отрезка кривой на кривой, площадь куска поверхности на поверхности. Эту меру мы будем обозначать m . В качестве \mathfrak{A} берется множество всех измеримых подмножеств Ω , то есть таких подмножеств, меру m которых можно вычислить. Геометрическая вероятность задается формулой

$$\mathbb{P}(A) = \frac{m(A)}{m(\Omega)}.$$

При выборе геометрической модели для вычисления вероятностей можно ориентироваться на инвариантность вероятности некоторой группы преобразований (например, вращений окружности или сферы, сдвигов на прямой, на плоскости, в пространстве). Если такая инвариантность напрашивается, то целесообразно применять геометрическую модель. Труднее обосновать применение геометрической модели, если Ω — отрезок (но не $[0, 2\pi]$), квадрат, прямоугольник или даже треугольник, круг, так как тогда такой группы преобразований нет. В задачах мы навязываем эту модель, утверждая, что точки соответствующего множества берутся наугад. На практике иногда приходится обращаться к геометрической модели в случае отрезка или квадрата, так как ничего лучшего у нас нет. Но в таком примере, как время ожидания автобуса или троллейбуса в условии заданного интервала движения (если расписание неизвестно), использование геометрической модели вполне оправдано. Иногда геометрическая модель оказывается теоретико-множественным произведением других более простых геометрических объектов (например, прямоугольник — произведение двух отрезков, в задаче на время ожидания двух автобусов время ожидание автобуса A и автобуса B — это пара чисел, если нас интересует событие 'суммарное время ожидания меньше какого-то x ', то эти числа нужно складывать). Геометрическая модель для распределения плоскостей, на которых врашаются девять планет вокруг Солнца — это произведение девяти сфер. Но на самом деле плоскости вращения планет очень близки друг к другу, в геометрической модели вероятность такой близости чрезвычайно мала. Таким образом, факт такой близости неслучаен и должен объясняться любой теорией возникновения Солнечной системы. Мы видим, что иногда геометрическая (да и классическая модель) строятся лишь для того, чтобы их отвергнуть и обосновать неслучайность некоторой наблюдаемой нами закономерности. (Астрономы утверждают, что близость плоскостей орбит всех планет Солнечной системой объясняется силой притяжения планеты Юпитер, притянувшей все планеты на одну и ту же плоскость эклиптики.)

Одним из примеров тщательного эксперимента с геометрической вероятностью является рулетка, если рулетка плохо вращается или останавливается чаще в одном месте, то игроки могут выигрывать у игорного дома.

Задача, в которой в условиях геометрической модели мы разными способами получаем разные (!) вероятности одного и того же события.

На окружности радиуса 1 наугад проводится хорда. Какова вероятность того, что ее длина больше $\sqrt{3}$?

Первый способ. Зафиксируем один из концов хорды, эта фиксация безразлична для нашей задачи, поэтому так мы можем сделать. Другой конец хорды пробегает всю окружность, и мы используем в качестве геометрической модели эту окружность. Мера точек окружности, для которых длина хорды больше $\sqrt{3}$, равна $2\pi/3$, поэтому вероятность равна $1/3$.

Заметим, что если мы один конец хорды не фиксируем, то в качестве Ω мы берем произведение двух окружностей.

Второй способ. Давайте зафиксируем направление хорды, т. е. перпендикулярный ей диаметр, который делит хорду пополам. Длина хорды будет больше $\sqrt{3}$, если точка пересечения будет лежать на половине диаметра, состоящего из точек, расстояние которых до центра меньше $1/2$. Тогда геометрическая модель — это точки диаметра, и наша вероятность равна $1/2$.

Третий способ. Давайте зафиксируем середину хорды и в качестве Ω будем рассматривать все точки внутри круга. Площадь таких точек равна π . Нашему событию соответствуют точки круга, отстоящие от центра не больше чем на $1/2$. Площадь множества этих точек равна $\pi/4$. Вероятность равна $1/4$.

Итак, разными способами вычисляя вероятность мы получаем разные ответы. По видимому, мы вычисляем вероятность одного и того же события, но получаемого в результате разной организации случайного эксперимента.

Все же первый способ представляется более разумным, так как он использует инвариантность вероятности относительно группы вращений окружности.

Замечание. Геометрическую вероятность на круге получить и представить в эксперименте значительно сложнее. Как бы мы не организовали эксперимент, будет очень трудно объяснить даже самим себе равенство вероятностей попадания в две фигуры в круге одинаковой площади, одна из которых близка к центру круга, а другая близка к краю. Впрочем, если моделировать геометрическую модель на квадрате, то ее можно смоделировать и на круге — использовать лишь те точки, попавшие на квадрат, которые попали также в круг. А вот как организовать эксперимент, моделью которого будет служить площадь на квадрате? Можно бросать с большой высоты точку на клетчатую бумагу, при этом результат каждого эксперимента переносить на выделенный квадрат. По существу, все квадраты клетчатой бумаги переносятся на один квадрат, благодаря этому бесконечная площадь бесконечной клетчатой бумаги становится конечной. Для моделирования геометрической модели на квадрате можно также использовать ту же идею, что и для круга — бросать точки с большой высоты, и брать во внимание, а следовательно, использовать при подсчете частоты, лишь те эксперименты, в которых точка попала на данный квадрат.

Задача 1. В метро в СПб интервал движения (время между приходом двух последовательных поездов) равен 4 минутам. Какова вероятность того, что ждать поезд придется меньше 1 мин.?

Решение. Используем геометрическую модель, в которой в качестве Ω используется интервал $(0, 4)$, $A = (0, 1)$. $\mathbb{P}(A) = \frac{m(0,1)}{m(0,4)} = 1/4$ (m — длина интервала, а в более общей ситуации — мера Лебега).

Задача 2. В метро в Москве интервал движения (время между приходом двух последовательных поездов) равен 2 минутам. Пассажир едет с одной пересадкой. Какова вероятность того, что суммарное время ожидания двух поездов будет меньше 1 мин.?

Решение. Опять используем геометрическую модель, но теперь ждать нужно два поезда, время ожидания первого поезда обозначим через x , второго — через y . Как x так и y попадают в интервал $(0, 1)$. Пара (x, y) попадает в квадрат $\{(x, y) : 0 < x < 1, 0 < y < 1\}$, который и используется в качестве Ω . $A = \{(x, y) : x+y < 1\} \cap \Omega$. $\mathbb{P}(A) = \frac{m(A)}{m(\Omega)} = 1/8$ (m — площадь).

Замечание. В случае метро использование геометрической модели наиболее оправдано. Для

автобуса, троллейбуса или трамвая интервал движения обычно расплывается. Тем не менее, использование геометрической модели оправдано, так как выработка адекватной вероятностной модели для данного маршрута потребовало бы слишком больших усилий.

Упражнение 4. В метро в Москве интервал движения (время между приходом двух последовательных поездов) равен 2 минутам. Пассажир едет с двумя пересадками. Какова вероятность того, что суммарное время ожидания всех трех поездов будет меньше 1.5 мин.?

§3. Вероятностные формулы. Условная вероятность

Мы рассмотрели два вида моделей — классические и геометрические. В первой модели вероятность события считается по формуле $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$, во второй — по формуле $\mathbb{P}(A) = \frac{m(A)}{m(\Omega)}$. Обоснованием в практической ситуации справедливости обеих моделей является соображение симметрии элементарных исходов. В классической модели симметрия осуществляется любой перестановкой множества всех элементарных исходов. Для применения этой модели нужно быть уверенным, что такая перестановка не меняет вероятность. Аналогичный подход возможен и при применении геометрической модели, но не всегда. Если множество всех исходов — точки окружности или сферы, то всегда существует вращение Ω , переводящее любую точку в любую. Для отрезка, квадрата, круга и шара это не так, и поэтому даже придумать ситуацию, в которой по всей видимости должна быть справедлива геометрическая модель, непросто. Например, если точка бросается на клетчатую бумагу, то мы предполагаем, что точки бросаются с большой высоты, используя симметричность точек при преобразовании не квадрата разбиения, а всей плоскости. Но так как мера на плоскости (площадь) бесконечна, то мы рассматриваем не точки плоскости, а классы эквивалентности точек, одинаково расположенных по отношению к клеткам клетчатой бумаги. Другими словами, все клетки клетчатой бумаги мы складываем на одну клетку, и для этой клетки используем геометрическую модель. При использовании геометрической модели для времени ожидания транспорта мы предполагаем наличие жестко заданного интервала движения, а случайному считаем время прихода на остановку (мы не знаем расписание). Впрочем, если отрезок превратить в окружность, то реализацией геометрической модели можно считать момент остановки быстро вращающейся рулетки. Но если посмотреть на игровую рулетку, то можно заметить, что при делении круга на девять равных областей, занумерованных числами от 1 до 9, каждая часть не состоит из одного угла, а является объединением углов в разных частях круга. Кроме удобства для игроков, это позволяет уменьшить ошибку от возможных искривлений рулетки.

Итак, если нет нужных симметрий, то мы пока не умеем вычислять вероятности, даже как функции некоторых параметров. Если мы бросаем монету и она не симметрична, то пока ничего не знаем о вероятности выпадения k гербов в n испытаниях, даже если вероятность выпадения герба равна 0.51 вместо нужных нам 0.5. Далее мы займемся формулами и понятиями, позволяющими иногда анализировать такие несимметричные ситуации.

Вероятностные формулы позволяют в рамках построенных классических или геометрических моделей облегчить вычисление вероятностей событий, сделать это вычисление более обоснованным. Знание этих формул поможет избежать элементарных ошибок. Часто применение таких формул позволяет одну громоздкую модель заменить на несколько более простых и более очевидных. Важнейшим понятием, которое будет введено в этом параграфе, является понятие условной вероятности.

1. Формула сложения.

Для двух событий формула имеет следующий вид:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB). \quad (1)$$

Доказательство очевидно. $A \cup B = A + (B \setminus A)$, поэтому

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A).$$

С другой стороны $B = (B \setminus A) + AB$, поэтому

$$\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(AB).$$

Эти равенства вместе дают (1).

Формула сложения для n событий

$$\begin{aligned} \mathbb{P}\left\{\bigcup_{i \leq n} A_i\right\} &= \sum_i \mathbb{P}(A_i) - \sum_{i < j} \mathbb{P}(A_i A_j) + \\ &+ \sum_{i < j < k} \mathbb{P}(A_i A_j A_k) - \dots + (-1)^{n+1} \mathbb{P}\left\{\prod_{i \leq n} A_i\right\}. \end{aligned} \quad (1')$$

Доказательство формулы сложения для произвольного n производится методом математической индукции. Предполагая доказанной формулу (1') для n слагаемых, в формуле для $n+1$ мы сначала рассматриваем $A_n \cup A_{n+1}$ как одно множество и применяем формулу для n , после чего используем в соответствующих слагаемых формулу сложения для $\mathbb{P}(A_n \cup A_{n+1})$ а также для вероятностей множеств вида

$$A_i A_j \cdots A_k (A_n \cup A_{n+1}) = A_i A_j \cdots A_k A_n \cup A_i A_j \cdots A_k A_{n+1},$$

используя, как мы только что сделали, дистрибутивность операций объединения и умножения (пересечения) событий.

Другой способ доказательства методом индукции состоит в использовании формулы

$$\mathbb{P}\left(\left\{\bigcup_{i \leq n} A_i\right\} \cup A_{n+1}\right) = \mathbb{P}\left\{\bigcup_{i \leq n} A_i\right\} + \mathbb{P}(A_{n+1}) - \mathbb{P}\left\{\bigcup_{i \leq n} A_i A_{n+1}\right\}.$$

Замечание. Иногда объединение бывает полезно представить как сумму несовместных событий и представить вероятность объединения как сумму вероятностей несовместных событий: например,

$$A_1 \cup \dots \cup A_n = A_1 + [A_2 \setminus A_1] + \dots + [A_n \setminus (A_1 \cup A_2 \cup \dots \cup A_{n-1})].$$

2. Условная вероятность. По определению *условная вероятность* события A при условии B вводится формулой

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}. \quad (2)$$

Разумность этого определения легко следует из частотной интерпретации вероятности. Мы рассматриваем не все эксперименты, а лишь те, в которых выполнилось условие, то есть произошло B , среди таких экспериментов берем лишь те, в которых произошло также A . В итоге

$$\frac{n_{AB}}{n_B} = \frac{n_{AB}}{n} / \frac{n_B}{n} \rightarrow \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}.$$

Разумеется, условная вероятность имеет смысл лишь при $\mathbb{P}(B) \neq 0$. Полезно четко понимать, что *условная вероятность* (при данном условии) – это вероятность. Действительно,

$$\mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega B)}{\mathbb{P}(B)} = 1, \mathbb{P}\left(\sum_n A_n | B\right) = \sum_n \mathbb{P}(A_n | B).$$

В частности, можно рассмотреть условную вероятность, в которой роль вероятности также играет условная вероятность. Что это означает?

$$\mathbb{P}\{(A|B)|C\} = \frac{\mathbb{P}(AC|B)}{\mathbb{P}(C|B)} = \frac{\mathbb{P}(ACB)}{\mathbb{P}(CB)} / \frac{\mathbb{P}(CB)}{\mathbb{P}(B)} = \frac{\mathbb{P}(ACB)}{\mathbb{P}(CB)} = \mathbb{P}(A|BC).$$

Комментарий. Понятие условной вероятности ввел Байес в середине 18 века. Первоначально это был набор понятий: рассматривались ситуации, когда событие A происходит после B , предшествует B и т.д. К математической теории эти подробности не имеют отношения. Мы вполне вправе рассматривать условную вероятность более раннего события A при условии выполнения более позднего события B , таким образом, мы в вероятностном смысле реконструируем более раннюю ситуацию по позднему результату.

Замечание. Бывают задачи, условия которых мы можем сформулировать в терминах условных вероятностей (см. ниже задачу о распаде атома). При переходе к условной вероятности $\mathbb{P}(A|B)$ мы меняем модель (заменяем множество всех элементарных исходов Ω на другое множество B , существенно разные события могут стать при такой замене одинаковыми, например, $\mathbb{P}(A|B) = \mathbb{P}(AB|B)$).

3. Формула умножения.

$$\mathbb{P}(A_1 A_2 \dots A_n) = \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \dots \mathbb{P}(A_n | A_1 \dots A_{n-1}). \quad (3)$$

Доказательство также производится методом математической индукции. Для двух сомножителей формула умножения следует из определения условной вероятности —

$$\mathbb{P}(AB) = \mathbb{P}(A|B)\mathbb{P}(B).$$

Для $n + 1$ событий имеем:

$$\begin{aligned} \mathbb{P}(A_1 A_2 \dots A_n A_{n+1}) &= \mathbb{P}(A_1 A_2 \dots A_n) \mathbb{P}(A_{n+1} | A_1 A_2 \dots A_n) = \\ &= \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \dots \mathbb{P}(A_n | A_1 \dots A_{n-1}) \mathbb{P}(A_{n+1} | A_1 A_2 \dots A_n). \end{aligned}$$

Замечание. Формула умножения чрезвычайно полезна, если мы можем как-то вычислять условную вероятность, но не можем без этой формулы вычислить вероятность произведения событий. В следующем примере формула умножения позволяет упростить модели.

Задача 1. Имеется 100 деталей, из них 90 — стандартные, 10 — бракованные. Выбираем 5 деталей (без возвращения). Какова вероятность того, что все они окажутся стандартными? Обозначим $A_i = \{i\text{-я деталь стандартна}\}$, тогда интересующее нас событие записывается как $A_1 A_2 A_3 A_4 A_5$. Вычисляем: $\mathbb{P}(A_1) = \frac{90}{100}$, условие A_1 означает, что у нас 99 деталей, из них 89 — стандартно, условие $A_1 A_2$ означает, что у нас 98 деталей, из них 88 — стандартно, и т. д. Таким образом, для вычисления каждой условной вероятности мы строим свою классическую модель, число элементарных исходов которой определяется соответствующим условием. Итак,

$$\mathbb{P}(A_1 A_2 \dots A_5) = \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \dots \mathbb{P}(A_5 | A_1 \dots A_4) = \frac{90}{100} \frac{89}{99} \frac{88}{98} \frac{87}{97} \frac{86}{96}.$$

Этот результат может быть получен и с помощью одной сложной модели, состоящих из упорядоченных (или неупорядоченных) выборок длины 5 без возвращений. Таким образом, в данной (и в последующей) задачах польза от применения формулы умножения незначительна.

4. Задача о рассеянном почтальоне. Почтальон должен был разнести n писем по n адресам. Но он перепутал все адреса случайным образом. Какова вероятность того, что хотя бы один адресат получил свое письмо?

Введем обозначение — $A_i = \{i\text{-е письмо дошло до своего адресата}\}$. Заодно мы нумеруем письма. Наше событие имеет вид $\bigcup_i A_i$. Для вычисления вероятности применим формулу сложения, для вычисления вероятностей произведений применим формулу умножения, а для вычисления условных вероятностей будем строить промежуточные классические модели. Имеем:

$$\mathbb{P}(A_i) = \frac{1}{n} \forall i, \mathbb{P}(A_i A_j) = \mathbb{P}(A_j | A_i) \mathbb{P}(A_i) = \frac{1}{n-1} \frac{1}{n} \forall i < j, \dots,$$

$$\mathbb{P}\left(\prod_{i \leq n} A_i\right) = \frac{1}{n!}$$

Далее мы учитываем то, что количество пар, троек и т.д. совпадает с соответствующим числом сочетаний, после очевидных выкладок получаем

$$\mathbb{P}\left(\bigcup_{i \leq n} A_i\right) = 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n+1} \frac{1}{n!} \approx 1 - e^{-1}.$$

5. Время жизни атома.

Как известно, атом в отличие от людей не стареет. Вероятность распада атома не зависит от того, сколько он до этого просуществовал. Если обозначить через ξ время жизни атома, то это обстоятельство можно записать следующей формулой:

$$\mathbb{P}\{\xi > t + h | \xi > t\} = \mathbb{P}\{\xi > h\}. \quad (*)$$

В итоге имеем:

$$\mathbb{P}\{\xi > t + h\} = \mathbb{P}[\{\xi > t + h\} \cap \{\xi > t\}] = \mathbb{P}\{\xi > t + h | \xi > t\} \mathbb{P}\{\xi > t\} = \mathbb{P}\{\xi > h\} \mathbb{P}\{\xi > t\}.$$

(Очевидно, что $\{\xi > t + h\} = \{\xi > t + h\} \cap \{\xi > t\}$.) Итак, вероятность для атома прожить дольше t удовлетворяет равенству, определяющему в случае непрерывности функции (см. 1 семестр, Анализ) экспоненциальную функцию. В нашем случае вместо непрерывности мы можем использовать убывание, и этого также достаточно. Итак,

$$\mathbb{P}\{\xi > t\} = e^{-\lambda t},$$

где $\lambda > 0$ для каждого вида атома свое. λ неотрицательно, так как в противном случае вероятность зашкалит за 1. Числовой множитель перед $e^{-\lambda t}$ равен $\mathbb{P}\{\xi > 0\} = 1$.

Замечание. Попробуем еще раз объяснить формулу (*). Если атом уже прожил t времени, то вероятность для него прожить еще h оказывается условной при условии $B_t = \{\text{атом прожил } t \text{ времени}\}$. То, что эта условная вероятность не зависит от t , отражает физические свойства атомов. Для других объектов, например, для человека, такое предположение неверно.

6. Формула полной вероятности.

Полной группой событий называется набор $\{A_i\}$ несовместных событий положительной вероятности, сумма которых равна Ω . Примером такого набора является множество всех элементарных исходов в классической модели. Каждый такой набор позволяет получить новую формулу для вычисления произвольного события B :

$$\mathbb{P}(B) = \sum_i \mathbb{P}(B|A_i) \mathbb{P}(A_i). \quad (4)$$

Доказательство очевидно: $\Omega = \sum_i A_i$, поэтому

$$\mathbb{P}(B) = \sum_i \mathbb{P}\left(B \sum_i A_i\right) = \sum_i \mathbb{P}(BA_i) = \sum_i \mathbb{P}(B|A_i) \mathbb{P}(A_i).$$

7. Формула Байеса.

Формула (4) вместе с определением условной вероятности и формулой умножения позволяет выразить условную вероятность $\mathbb{P}(A_i|B)$ через $\mathbb{P}(B|A_i)$. А именно,

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i) \mathbb{P}(A_i)}{\sum_i \mathbb{P}(B|A_i) \mathbb{P}(A_i)}. \quad (5)$$

8. Задача о разорении игрока.

Два игрока играют одинаково сильно (т. е. каждый из них выигрывает одну партию с вероятностью $1/2$) до разорения одного из них. Стоимость партии равна 1 рублю. У игрока A — m рублей, у игрока B — n рублей. Какова вероятность того, что разорится A ? Эту задачу можно трактовать также как задачу о выигрыше B . Мы так и будем делать, сохранив, тем не менее, за задачей общепринятое название.

Заметим, что игра может продолжаться сколь угодно и поэтому множество всех элементарных исходов бесконечно. Это затрудняет использование классической модели даже несмотря на симметрию ситуации.

Пожалуй, идея решения заключается в введении обозначения для общей суммы денег — $s = m+n$, тогда $m = s-n$. Теперь изменение числа n рублей у игрока B не меняет общую сумму. Обозначим через $p(n)$ вероятность окончательного выигрыша B , если у B на данный момент n рублей. После первой партии у B будет либо $n+1$ либо $n-1$ рублей, причем вероятность и того и другого равна $1/2$. Разложим вероятность окончательного выигрыша B по формуле полной вероятности по полной группе событий

$$\Omega = \{\text{выигрыш 1 партии}\} + \{\text{проигрыш 1 партии}\}.$$

Имеем:

$$p(n) = \frac{1}{2}p(n-1) + \frac{1}{2}p(n+1) \quad \forall n, p(s) = 1, p(0) = 0.$$

Нетрудно догадаться, что этой системе равенств (для всех n от нуля до s) удовлетворяет функция $p(n) = \frac{n}{s}$.

Для доказательства единственности этого решения надо вспомнить линейную алгебру. Мы имеем $s+1$ равенств:

$$\begin{aligned} p(0) &= 0; \\ p(1) &= \frac{1}{2}p(0) + \frac{1}{2}p(2) \\ &\dots \\ p(s-1) &= \frac{1}{2}p(s-2) + \frac{1}{2}p(s); \\ p(s) &= 1. \end{aligned}$$

В этой системе лишь $s-1$ равенств содержат неизвестные $p(n)$. Определитель системы равен

$$\left| \begin{array}{cccccc} -1 & 1/2 & 0 & 0 & \dots & 0 \\ 1/2 & -1 & 1/2 & 0 & \dots & 0 \\ 0 & 1/2 & -1 & 1/2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 1/2 & -1 & 1/2 \\ 0 & \dots & \dots & 0 & 1/2 & -1 \end{array} \right|.$$

Нетрудно убедиться в том, что определитель системы не равен нулю, а это означает, что решение $p(n) = \frac{n}{s}$ является единственным. Игра может длиться сколь угодно, и число $p(n)$ можно представить в виде суммы ряда. Студенты-механики задали мне очень существенный вопрос: а не может ли так оказаться, что вероятность события C , которое означает, что никто не выиграет за бесконечное время, не равна нулю? Из нашей формулы следует, что это невозможно:

$$p(n) + p(m) = \frac{n}{n+m} + \frac{m}{n+m} = 1.$$

Но это можно доказать и по другому. При любом числе игр k вероятность C оценивается (очень грубо) сверху суммой вероятностей $\alpha(n)$ того, что на данный момент игрок A выиграл на n

партий больше игрока B , где n пробегает все числа от $-s$ до s . Легко проверить (см. раздел 'Случайное блуждание' и формулу Стирлинга), что

$$\begin{aligned}\alpha(n) &= C_k^{(n+k)/2} 2^{-k} = \frac{k!}{[(k+n)/2]![(k-n)/2]!} \leq \alpha(0) \approx \\ &\approx \left(\frac{k}{e}\right)^k 2\sqrt{\pi k} / \left[\left(\frac{k/2}{e}\right)^{k/2} 2\sqrt{\pi k/2} \left(\frac{k/2}{e}\right)^{k/2} 2\sqrt{\pi k/2} \right] 2^{-k} = \frac{1}{\sqrt{\pi k}} \rightarrow 0.\end{aligned}$$

Поэтому и вероятность C оценивается числом $\frac{2s}{\sqrt{\pi k}}$ (s фиксировано) и сходится к 0 при $k \rightarrow \infty$.

Упражнение. Решите аналогичную задачу в ситуации, когда вероятность выигрыша равна p , а проигрыша $1 - p$. [Указание. Используйте формулу для геометрической прогрессии.]

9. Разрушающий контроль качества.

Формула Байеса позволяет использовать для контроля качества априорную информацию — производственный опыт.

Допустим, что для проверки качества изделия его нужно разрушить. В этой ситуации используется контроль качества не одного изделия, а партии изделий.

Пример. Допустим, что на заводе производятся электрические лампы. В нормальной ситуации производятся стандартные партии ламп, в которых 90 процентов ламп стандартно, а 10 процентов ламп браковано. Обычно (с вероятностью 0.91) завод выпускает стандартные партии. Однако иногда производство разлаживается (с вероятностью 0.09) и выпускаются бракованные партии, в которых 50 процентов ламп стандартно, а 50 процентов ламп браковано. Выборочный контроль качества партии осуществляется следующим образом — испытывается (и разрушается) одна лампа, если она стандартна, то партия объявляется стандартной, если она бракована, то партия объявляется бракованной. Насколько эффективен и разумен такой контроль?

Без контроля согласно производственному опыту доля бракованных партий равна 0.09. Посмотрим, какова доля бракованных партий среди объявленных стандартными. Введем события:

$$B = \{\text{выбранная лампа стандартна}\}, A_1 = \{\text{партия стандартная}\}, A_2 = \{\text{партия бракованная}\}.$$

Интересующее нас событие имеет вид $\mathbb{P}(A_2|B)$. Считаем:

$$\mathbb{P}(A_2|B) = \frac{\mathbb{P}(B|A_2)\mathbb{P}(A_2)}{\mathbb{P}(B|A_1)\mathbb{P}(A_1) + \mathbb{P}(B|A_2)\mathbb{P}(A_2)} = \frac{0.5 \cdot 0.09}{0.5 \cdot 0.09 + 0.9 \cdot 0.91}.$$

В результате нашего контроля доля бракованных партий среди принятых существенно понизилась.

10. Урновая модель.

Довольно много задач используют так называемую урновую модель.

Пример. В комнате находятся две урны. Известно, что в первой урне 4 белых и 2 черных шара, а во второй — 2 белых и 3 черных шара. Вы зашли в комнату, наугад выбрали урну, вытащили из нее шар, шар оказался белым. Какова вероятность того, что вы вытащили шар из первой урны?

Казалось бы ответ очевиден: урну выбрали наугад, они обе равноправны, следовательно, вероятность вытащить шар из каждой урны равна $1/2$. Но давайте рассмотрим другую, более определенную ситуацию — во второй урне нет черных шаров. Тогда из цвета вытащенного шара немедленно следует, что он вынут из первой урны, т.е. вероятность первой урны равна 1. Итак, результат эксперимента меняет вероятность обращения к каждой из урн. Введем обозначения: $B = \{\text{вытащили белый шар}\}$, $A_i = \{\text{обратились к } i\text{-й урне}\}$. В начальный момент вероятность обращения к каждой урне равна $1/2$,

$$\mathbb{P}(B) = \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \mathbb{P}(B|A_2)\mathbb{P}(A_2).$$

Как мы уже отметили, $\mathbb{P}(A_1) = \mathbb{P}(A_2) = 1/2$. Из известного нам состава урн следует, $\mathbb{P}(B|A_1) = 4/6$, $\mathbb{P}(B|A_2) = 2/5$, поэтому

$$\mathbb{P}(B) = \frac{4}{6} \cdot \frac{1}{2} + \frac{2}{5} \cdot \frac{1}{2} = \frac{8}{15}.$$

Теперь мы можем подсчитать изменившиеся в результате эксперимента вероятности каждой из урн по формуле Байеса:

$$\mathbb{P}(A_1|B) = \frac{\mathbb{P}(A_1B)}{\mathbb{P}(B)} = \frac{1/3}{8/15} = \frac{5}{8}, \mathbb{P}(A_2|B) = \frac{3}{8}.$$

Давайте решим еще одну задачу. После того, как мы вытащили белый шар из урны, мы поменяли урну и вытащили из шар из другой урны. Какова вероятность того, что он оказался белым?

Решая эту задачу, мы должны использовать уже измененные вероятности урн. Обозначим второе событие через B' . Легко видеть, что нас интересует $\mathbb{P}(B'|B)$. Итак, мы используем измененную формулу полной вероятности, в которой все вероятности оказываются условными, причем вероятность того, что другая урна является первой, равна теперь $3/8$, а вероятность того, что она является второй, равна $5/8$:

$$\mathbb{P}(B'|B) = \mathbb{P}(B'|A_1)\mathbb{P}(A_1|B) + \mathbb{P}(B'|A_2)\mathbb{P}(A_2|B) = \frac{3}{8} \cdot \frac{2}{3} + \frac{5}{8} \cdot \frac{2}{5} = \frac{1}{2}.$$

Упражнение 5. В условиях задачи 2 §2 найти условную вероятность

$$\mathbb{P}(\text{время ожидания 1-го поезда } < 0.5 \text{ мин.} | \text{суммарное время ожидания } < 1 \text{ мин.}).$$

§4. Независимость

К определению независимости можно прийти из определения условной вероятности. Естественно назвать событие A независимым от B , если $\mathbb{P}(A|B) = \mathbb{P}(A)$. Умножив равенство на $\mathbb{P}(B)$, мы получаем:

1. **Определение.** События A и B называются *независимыми*, если

$$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B).$$

Это определение не только симметрично, но и имеет смысл также и для невозможного события \emptyset .

Очевидно, что *невозможное событие \emptyset и достоверное событие Ω независимы от любого другого события*.

Примеры независимых событий можно привести для классической и геометрической моделей. Например, когда в качестве модели мы используем множество всех упорядоченных выборок с возвращениями из множества Ω из n элементов, то события

$$A = \{\text{1-й элемент выборки есть данное } \omega_1\},$$

$$B = \{\text{2-й элемент выборки есть данное } \omega_2\}$$

независимы. Проверим:

$$\mathbb{P}(AB) = n^{n-2}/n^n = 1/n \cdot 1/n = \mathbb{P}(A)\mathbb{P}(B).$$

Но независимости нет, если в качестве Ω используется множество всех выборок без возвращений (разумеется, тогда $\omega_1 \neq \omega_2$): в этом случае

$$\mathbb{P}(A) = \mathbb{P}(B) = 1/n, \mathbb{P}(AB) = 1/n(n-1).$$

Все же при больших n события A и B 'примерно' независимы. Аналогичный пример независимости можно придумать для геометрической модели. Пусть Ω — квадрат $[0, 1] \times [0, 1]$, вероятность — обычная площадь. Тогда события

$$A = \{(x, y) \in \Omega : x \in (a, b)\} \text{ и } B = \{(x, y) \in \Omega : y \in (c, d)\}$$

независимы:

$$\begin{aligned}\mathbb{P}(A) &= b - a, \mathbb{P}(B) = d - c, \\ \mathbb{P}(AB) &= \mathbb{P}\{(x, y) : x \in (a, b), y \in (c, d)\} = (b - a)(d - c).\end{aligned}$$

Ситуация изменится, если в качестве модели Ω будет использоваться не квадрат, а круг $\Omega = \{(x, y) : x^2 + y^2 \leq 1\}$ с вероятностью \mathbb{P} — площадью, деленной на π . Тогда для введенных нами множеств A и B равенство $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$ будет неверно. Проверьте сами, что при

$$A = \{(x, y) \in \Omega : x \in (-\sqrt{1/2}, \sqrt{1/2})\},$$

$$B = \{(x, y) \in \Omega : y \in (-\sqrt{1/2}, \sqrt{1/2})\}$$

имеют место равенства $\mathbb{P}(A) = \mathbb{P}(B) = (\pi/2 + 1)/\pi$, $\mathbb{P}(AB) = 1/\pi$. При a, b, c, d вблизи 0 события A и B будут также зависимы и число $\mathbb{P}(AB)/\mathbb{P}(A)\mathbb{P}(B)$ будет примерно равно $\pi/4$.

Пример: Независимы пары событий 'пики', 'маленькие' и 'трэфи', 'маленькие', которые в сумме дают независимую пару 'черные', 'маленькие'. Независимость этих пар событий имеет место как для колоды из 36 карт, так и для колоды из 52 карт. Но если добавить к этой колоде один джокер, то пример окажется несостоительным. Более того, я рекомендую разобраться, почему в классической модели из 53 карт нет нетривиальных (т. е. не содержащих достоверное или невозможное событие) пар независимых событий.

2. Условная независимость.

События A и B называются *условно независимыми* относительно события C , если

$$\mathbb{P}(AB|C) = \mathbb{P}(A|C)\mathbb{P}(B|C).$$

Приведенные выше события в круге условно независимы относительно события $C = (c, d) \times (e, f)$, заданного любым прямоугольником внутри круга, стороны которого параллельны осям координат.

Упражнения. Пусть $A \cup B \subset C$, $0 < \mathbb{P}(C) < 1$. Покажите, что два условия 'события A и B независимы' и 'события A и B условно независимы при условии C' ' одновременно выполняться не могут. Пусть события A , B и C независимы в совокупности (определение см. ниже). Проверьте, что тогда A и B условно независимы при условии C . Могут ли события A и B быть условно независимыми относительно события $A \cup B$? события AB ?

3. Несколько эквивалентных определений независимости в совокупности.

Теорема. Следующие определения эквивалентны:

а) События A_1, A_2, \dots, A_n называются *независимыми в совокупности*, если для любого множества индексов $I \subset \overline{1, n}$ имеет место равенство:

$$\mathbb{P}\left(\prod_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i). \quad (0).$$

б) События A_1, A_2, \dots, A_n называются *независимыми в совокупности*, если для любых непересекающихся множеств индексов $I \subset \overline{1, n}$ и $J \subset \overline{1, n}$ события $\prod_{i \in I} A_i$ и $\prod_{i \in J} A_i$ независимы.

с) События A_1, A_2, \dots, A_n называются *независимыми в совокупности*, если для любого набора событий B_1, B_2, \dots, B_n , где каждое B_i равно или A_i или $\overline{A_i}$, выполняется равенство

$$\mathbb{P}\left(\prod_{i \leq n} B_i\right) = \prod_{i \leq n} \mathbb{P}(B_i). \quad (1)$$

Доказательство. а) \Rightarrow б) очевидно.

$$\begin{aligned} \mathbb{P}\left(\prod_{i \in I} A_i\right) \mathbb{P}\left(\prod_{i \in J} A_i\right) &= \prod_{i \in I} \mathbb{P}(A_i) \prod_{i \in J} \mathbb{P}(A_i) = \\ &= \prod_{i \in I+J} \mathbb{P}(A_i) = \mathbb{P}\left(\prod_{i \in I+J} A_i\right) = \mathbb{P}\left(\prod_I A_i \prod_J A_i\right). \end{aligned}$$

б) \Rightarrow а) доказывается индукцией по числу элементов в I . Например,

$$\mathbb{P}(A_i A_j A_k) = \mathbb{P}(A_i) \mathbb{P}(A_j A_k) = \mathbb{P}(A_i) \mathbb{P}(A_j) \mathbb{P}(A_k).$$

Аналогично, случай $|I| = 4$ сводится к случаю $|I| = 3$ и т. д.

а) и б) эквивалентны даже более сильному варианту с):

с') где (1) заменяется на

$$\mathbb{P}\left(\prod_{i \in I} B_i\right) = \prod_{i \in I} \mathbb{P}(B_i). \quad (1')$$

для любого множества индексов $I \subset \overline{1, n}$. По видимому, доказательство эквивалентности а) и с) проще всего провести через эквивалентность с с'), как это делается ниже.

Доказательство а) \Rightarrow с') ведется индукцией по числу элементов j в I , для которых B_j равно \overline{A}_j (для всех других $i \in I$ имеет место $B_i = A_i$). Покажем, как из (0) выводится (1'), если лишь для одного $j \in I$ имеет место $B_j = \overline{A}_j$. Мы рассмотрим два равенства (0) с множествами индексов $I \setminus \{j\}$ и I .

$$\begin{aligned} \mathbb{P}\left(\prod_{i \in I \setminus \{j\}} A_i\right) &= \prod_{i \in I \setminus \{j\}} \mathbb{P}(A_i). \\ \mathbb{P}\left(\prod_{i \in I} A_i\right) &= \prod_{i \in I} \mathbb{P}(A_i). \end{aligned}$$

Вычитая из верхнего равенства нижнее получаем:

$$\mathbb{P}\left(\prod_{i \in I \setminus \{j\}} A_i \overline{A}_j\right) = \prod_{i \in I \setminus \{j\}} \mathbb{P}(A_i) \mathbb{P}(\overline{A}_j).$$

Импликация с') \Rightarrow а) очевидна. Поэтому нам достаточно доказать с) \Rightarrow с'). Доказательство индукцией по числу сомножителей, но от n к 2, то есть на каждом следующем шаге число сомножителей уменьшается. Допустим, что мы уже доказали (1') для $k + 1$ сомножителей, докажем его для k сомножителей. Выберем индекс $j \notin I$. Согласно предположению верны равенства

$$\begin{aligned} \mathbb{P}\left(\prod_{i \in I} B_i A_j\right) &= \prod_{i \in I} \mathbb{P}(B_i) \mathbb{P}(A_j), \\ \mathbb{P}\left(\prod_{i \in I} B_i \overline{A}_j\right) &= \prod_{i \in I} \mathbb{P}(B_i) \mathbb{P}(\overline{A}_j). \end{aligned}$$

Складывая эти равенства, получаем

$$\mathbb{P}\left(\prod_{i \in I} B_i\right) = \prod_{i \in I} \mathbb{P}(B_i).$$

Замечание. Рекомендуется построить пример зависимых событий A_1, A_2, \dots, A_n таких, что выполнено одно равенство

$$\mathbb{P}\left(\prod_{i \leq n} A_i\right) = \prod_{i \leq n} \mathbb{P}(A_i).$$

4. Модели, предполагающие независимость в совокупности.

Пример с электрической цепью. Рассматривается электрическая цепь, в которую включено некоторое число разных сопротивлений. Надо найти вероятность прохода тока (хотя бы по одной из линий) от начала цепи до конца. Мы предполагаем независимость в совокупности событий, состоящих в исправности разных сопротивлений. В вычислениях используется формула сложения вероятностей (в случае параллельного расположения сопротивлений) и формула произведения (в случае последовательного расположения).

Упражнение. Симметричная монета. Классическая модель (любой набор длины n из гербов или решек имеет вероятность 2^{-n}) означает, что события 'герб' (или решка) в i испытании при разных i независимы в совокупности. Проверьте.

5. Пример Бернштейна попарно независимых событий, которые не являются независимыми в совокупности.

Рассматривается правильный тетраэдр, который с равной вероятностью может упасть на любую из своих граней. Покрасим боковые стороны тетраэдра в красный, синий и зеленый цвета, а основание во все три цвета. События K, S, Z имеют место, если на грани тетраэдра, на которую он упал, имеется соответствующий цвет. Непосредственно проверяется, что события K, S, Z попарно независимы, в то же время

$$\mathbb{P}(KSZ) = 1/4 \neq 1/8 = \mathbb{P}(K)\mathbb{P}(S)\mathbb{P}(Z).$$

6. Независимые σ -алгебры.

Определение. σ -подалгебры \mathfrak{A}_i ($i \leq n$) событий σ -алгебры \mathfrak{A}_i называются независимыми, если для любых событий $A_i \in \mathfrak{A}_i$ имеет место равенство

$$\mathbb{P}\left(\prod_i A_i\right) = \prod_i \mathbb{P}(A_i).$$

Любопытно, что добавка 'в совокупности' оказывается в этом случае излишней, так как из этого определения следует независимость в совокупности событий A_i . Мы требуем выполнения лишь одного равенства, но для всех наборов, состоящих из одного события в каждой из σ -алгебр независимость в совокупности следует из определения с). Действительно, каждое событие в наборе можно заменить на его отрицание.

Замечания. 1) Как конструировать примеры независимых σ -алгебр? Для этого можно рассмотреть последовательность или конечное число вероятностных пространств $(\Omega_i, \mathfrak{A}_i, \mathbb{P}_i)$ и взять теоретико-множественное произведение $\prod_i \Omega_i$, а в нем рассмотреть σ -алгебру \mathfrak{A} , порожденную всевозможными множествами вида $\prod_i A_i$, где $A_i \in \mathfrak{A}_i$. Можно доказать, но это не просто, и мы этого делать не будем (это аналог известной вам теоремы Фубини), что функция множеств

$$\mathbb{P}\left(\prod_i A_i\right) = \prod_i \mathbb{P}_i(A_i) \tag{*}$$

продолжается до вероятностной меры на всю σ -алгебру \mathfrak{A} . Первоначальным σ -алгебрам \mathfrak{A}_i соответствуют (мы их будем обозначать так же) σ -подалгебры σ -алгебры \mathfrak{A} , состоящие из множеств вида $\prod_i B_i$, где B_j равно Ω_j при $j \neq i$ и B_i равно A_i , а A_i пробегает \mathfrak{A}_i . Из равенства (*) следует независимость этих σ -подалгебр \mathfrak{A}_i .

Теперь мы сможем смоделировать ситуацию (см. Введение) — проводится последовательность экспериментов в одинаковых условиях. Один эксперимент моделируется вероятностным пространством $(\Omega, \mathcal{A}, \mathbb{P})$, далее берется степень Ω^N , где N — множество натуральных чисел, осуществляющая для этого бесконечного произведения указанную выше конструкцию, а в качестве события A , наблюдаемого в n -м эксперименте, возьмем событие $\prod_i B_i$, где B_i равно Ω при $i \neq n$ и B_n равно A .

Заметим, что удобное для примеров вероятностное пространство — отрезок $[0, 1]$ с σ -алгеброй борелевских множеств и мерой Лебега изоморфен счетной степени вероятностного пространства $\{0, 1\}$, где $\mathbb{P}\{1\} = \mathbb{P}\{0\} = 1/2$.

7. Операции с независимыми событиями.

Для дальнейшего представляет интерес вопрос — можно ли с помощью теоретико-множественных операций увеличивать число событий, независимых в совокупности. Например, пусть нам даны полуалгебры \mathfrak{C}_i , $i \leq n$, и все наборы событий $A_i \in \mathfrak{C}_i$, $i \leq n$, независимы в совокупности. Будут ли независимы в совокупности все наборы $A_i \in \mathfrak{U}_i$, $i \leq n$, где \mathfrak{U}_i — порожденная \mathfrak{C}_i , $i \leq n$, алгебра? Мы докажем это утверждение, и более того, докажем в этой ситуации независимость σ -алгебр \mathfrak{A}_i , порожденных \mathfrak{C}_i . Для упрощения обозначений мы рассмотрим случай лишь двух полуалгебр, \mathfrak{C}_1 и \mathfrak{C}_2 .

Напомним определение *полуалгебры* — это полукольцо, содержащее Ω , то есть набор множеств, замкнутый относительно конечных пересечений, и такой, что из принадлежности к нему двух множеств A и B следует, что множество $A \setminus B$ представимо в виде конечной суммы попарно непересекающихся элементов полуалгебры. Полуалгебра содержит и все пространство Ω (пересечение пустого семейства элементов полуалгебры). Существенным примером полуалгебры является набор всех полуоткрытых интервалов вида $[a, b)$ на числовой прямой. Другой существенный пример — это произведение алгебр (или полуалгебр). Если \mathfrak{C}_1 — полуалгебра в Ω_1 , а \mathfrak{C}_2 — полуалгебра в Ω_2 , то всевозможные произведения $A_1 \times A_2$, где A_1 пробегает \mathfrak{C}_1 , A_2 пробегает \mathfrak{C}_2 , образуют полуалгебру в $\Omega_1 \times \Omega_2$. Это утверждение легко проверяется и верно, разумеется, для любого конечного числа сомножителей. Тем более оно верно, если A_1 и A_2 берутся из двух σ -алгебр.

Мы приведем несколько правил об операциях с парами независимых событий, которые сохраняют независимость.

1) Если события A, B независимы, то события A, \bar{B} тоже независимы. Действительно, мы вычитаем из

$$\mathbb{P}(A\Omega) = \mathbb{P}(A) \cdot 1$$

равенство

$$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$$

и получаем

$$\mathbb{P}(A\bar{B}) = \mathbb{P}(A)\mathbb{P}(\bar{B}).$$

Используя этот факт, мы получаем, что независимость A, B означает независимость четырех пар (A, B) , (A, \bar{B}) , (\bar{A}, B) , (\bar{A}, \bar{B}) .

2) Если $A_1 A_2 = \emptyset$, а пары (A_1, B) , (A_2, B) независимы, то пара $(A_1 + A_2, B)$ независима. Доказательство сложением равенств

$$\mathbb{P}(A_1 B) = \mathbb{P}(A_1)\mathbb{P}(B),$$

$$\mathbb{P}(A_2 B) = \mathbb{P}(A_2)\mathbb{P}(B).$$

Замечание. Если убрать условие несовместности, то все не так: события $(K \cup S, Z)$ в примере Бернштейна зависимы. Однако ниже мы увидим, что из независимости в совокупности событий A, B, C следует независимость событий $(A \cup B, C)$.

Все сказанное в пунктах 1), 2) верно и для наборов событий, независимых в совокупности.

Вспомним, что порожденная полуалгеброй алгебра состоит из всевозможных конечных сумм попарно непересекающихся элементов полуалгебры. Таким образом, с учетом возможностей, предоставляемых методом математической индукции, из 1) и 2) следует:

8. Предложение. Пусть \mathfrak{F}_1 и \mathfrak{F}_2 — независимые полуалгебры. Тогда порожденные ими алгебры \mathfrak{U}_1 и \mathfrak{U}_2 также независимы.

Теперь рассмотрим операцию предела последовательности событий. Но предварительно докажем свойство вероятностной меры, которое называется *непрерывностью*.

9. Определение. Пусть последовательность событий (A_n) и событие A таковы, что $A_1 \subset A_2 \subset \dots \subset A_n \dots$ и $A = \bigcup_n A_n$ или $A_1 \supset A_2 \supset \dots \supset A_n \dots$ и $A = \bigcap_n A_n$. В каждом из этих случаев мы будем говорить, что последовательность (A_n) имеет предел A и обозначать $A_n \rightarrow A$, для удобства мы будем также вводить дополнительную нагрузку на символ и писать в первом случае $A_n \uparrow A$, а во втором случае $A_n \downarrow A$.

10. Определение. Функция множеств μ называется *непрерывной*, если $A_n \rightarrow A$ влечет $\mu(A_n) \rightarrow \mu(A)$.

11. Предложение. Вероятностная мера \mathbb{P} на σ -алгебре непрерывна.

Доказательство. Пусть $A_n \uparrow A$. Тогда

$$A = A_1 + (A_2 \setminus A_1) + \dots$$

$$A_n = A_1 + (A_2 \setminus A_1) + \dots + (A_n \setminus A_{n-1}).$$

В силу конечной аддитивности μ

$$\mathbb{P}(A_n) = \mathbb{P}(A_1) + \mathbb{P}(A_2 \setminus A_1) + \dots + \mathbb{P}(A_n \setminus A_{n-1})$$

и в силу счетной аддитивности

$$\mathbb{P}(A) = \mathbb{P}(A_1) + \mathbb{P}(A_2 \setminus A_1) + \dots.$$

Поэтому, так как сумма ряда по определению — предел частичных сумм ряда, мы имеем:

$$\mathbb{P}(A) = \lim_n \mathbb{P}(A_n).$$

Непрерывность для последовательности $A_n \downarrow A$ выводится из непрерывности для $A_n \uparrow A$ переходом к дополнениям.

12. Предложение. Если пары событий (A_n, B) независимы для всех n , $A_n \uparrow A$ или $A_n \downarrow A$, то события A и B также независимы. Доказательство переходом к пределу в определении независимости.

Далее мы покажем, что из независимости алгебр следует независимость порожденных ими σ -алгебр. Для этого нам надо показать, что порожденная σ -алгебра получается из алгебры лишь (сохраняющими независимость) переходами к пределу возрастающих или убывающих последовательностей. В теории меры имеется понятие, использующееся для исследования этого процесса.

13. Определение. Монотонным классом называется класс множеств (в каком-то множестве, содержащем все множества из класса, которое нам будет удобно по-прежнему обозначать Ω), замкнутый относительно перехода к пределу возрастающей или убывающей последовательности множеств (то есть вместе с любой такой последовательностью этот класс содержит и ее предел).

14. Предложение. Для любого набора множеств \mathfrak{F} существует наименьший содержащий \mathfrak{F} монотонный класс.

Доказательство очевидно и применяется для доказательства многих других утверждений такого рода. 1) Монотонные классы, содержащие \mathfrak{F} , существуют, например, множество всех

подмножеств Ω . 2) Рассмотрим все монотонные классы, содержащие \mathfrak{F} , и возьмем их пересечение. 3) Легко проверяется (??), что это пересечение является наименьшим монотонным классом.

В то же время, в более или менее сложной ситуации выписать явно все элементы наименьшего монотонного класса невозможно. Постараюсь дать представление о том, что получается. Сначала мы добавим к \mathfrak{F} все пределы монотонных последовательностей из \mathfrak{F} , получим \mathfrak{F}_1 . Но в \mathfrak{F}_1 появляются новые монотонные последовательности, добавив их пределы к \mathfrak{F}_1 , получим \mathfrak{F}_2 . Продолжим эту процедуру счетное число раз, получим возрастающую последовательность множеств (\mathfrak{F}_n) . Объединение этой последовательности обозначается \mathfrak{F}_ω . Но и это объединение не является монотонным классом, в нем появятся монотонные последовательности множеств по одному из каждого \mathfrak{F}_n . Добавив их к \mathfrak{F}_ω , получим $\mathfrak{F}_{\omega+1}$. Этот процесс называется трансфинитным. В течение процесса мы получим множества вида \mathfrak{F}_ω , $\mathfrak{F}_{\omega\omega}$, $\mathfrak{F}_{\omega\omega\omega}$ и т. д., вплоть до первого несчетного трансфинитного числа.

15. Предложение. σ -алгебра — это алгебра, которая является монотонным классом.

Доказательство в обе стороны очевидно. То, что σ -алгебра является монотонным классом, следует из того, что предел монотонной последовательности множеств на деле является объединением или пересечением. Обратно, любое счетное объединение (или пересечение) является монотонным пределом конечных объединений (или пересечений).

16. Теорема (о монотонных классах). *Наименьший монотонный класс \mathfrak{M} , содержащий алгебру \mathfrak{A} , является σ -алгеброй.*

Доказательство этой теоремы нетривиально. Очевидно, что \mathfrak{M} также содержит Ω и \emptyset , как и \mathfrak{A} . Просто доказывается, что \mathfrak{M} замкнут относительно операции перехода к дополнению. Мы рассматриваем вместе с \mathfrak{M} класс \mathfrak{M}^c , состоящий из дополнений всех множеств из \mathfrak{M} . По построению \mathfrak{M}^c также является монотонным классом (если $A_n \uparrow A$ и $A_n \in \mathfrak{M}^c$, то $A_n^c \downarrow A^c$ в \mathfrak{M} и $A = (A^c)^c \in \mathfrak{M}^c$). Но тогда $\mathfrak{M} \cap \mathfrak{M}^c$ — также монотонный класс (почему, проверьте хотя бы здесь!) меньший \mathfrak{M} .

Итак, нам достаточно доказать, что \mathfrak{M} замкнут относительно операций объединения и пересечения конечного числа множеств, причем достаточно это доказать для двух множеств.

1) Введем класс \mathfrak{M}_0 всех множеств $B \subset \Omega$ со свойством: для любого $A \in \mathfrak{A}$ множества $A \cup B$ и $A \cap B$ принадлежат \mathfrak{M} . Проверим, что \mathfrak{M}_0 — монотонный класс. Действительно, если $B_n \downarrow B$, $A \cup B_n$ и $A \cap B_n$ для любого n принадлежат \mathfrak{M} , то так как \mathfrak{M} — монотонный класс, $A \cup B = \lim_n (A \cup B_n)$ и $A \cap B = \lim_n (A \cap B_n)$ принадлежат \mathfrak{M} . Итак, \mathfrak{M}_0 — монотонный класс. Так как \mathfrak{A} — алгебра, элементы $B \in \mathfrak{A}$ также принадлежат \mathfrak{M}_0 . По определению \mathfrak{M} , \mathfrak{M}_0 содержит \mathfrak{M} . Следовательно, для любых $A \in \mathfrak{A}$, $B \in \mathfrak{M}$ тем более имеет место $A \cup B$, $A \cap B \in \mathfrak{M}$.

2) Введем класс \mathfrak{M}_{00} всех множеств $B \subset \Omega$ со свойством: для любого $A \in \mathfrak{M}$ множества $A \cup B$ и $A \cap B$ принадлежат \mathfrak{M} . Как и выше проверяется (??), что \mathfrak{M}_{00} — монотонный класс. Теперь проверим, что \mathfrak{M}_{00} содержит \mathfrak{A} . Действительно, в первом пункте показано, что для любого $A \in \mathfrak{A}$ и $B \in \mathfrak{M}$ множества $A \cup B$ и $A \cap B$ принадлежат \mathfrak{M} . Поменяв местами обозначения A и B , мы показываем, что все элементы $A \in \mathfrak{A}$ имеют свойство, определяющее \mathfrak{M}_{00} .

Итак, \mathfrak{M}_{00} — монотонный класс, \mathfrak{M}_{00} содержит \mathfrak{A} , значит \mathfrak{M}_{00} содержит \mathfrak{M} . В итоге для любых элементов $A \in \mathfrak{M}$, $B \in \mathfrak{M}$, элементы $A \cup B$ и $A \cap B$ принадлежат \mathfrak{M} .

17. Несимметричная модель Бернулли (биномиальная модель).

Несимметричной монета подбрасывается n раз. В каждом эксперименте вероятность герба равна p , а вероятность решки — $1 - p$. Как и в симметричной модели Бернулли, в качестве вероятностного пространства служит множество всех наборов из символов Г и Р длины n . Гербу в i -испытании соответствует событие G_i , состоящее из всех наборов гербов и решек, в которых на i -м месте находится герб, так же определяется событие R_i . Так как герб или решка в одном испытании не влияют на условия проведения другого испытания, то все σ -алгебры $\mathfrak{A}_i = \{\Omega, G_i, R_i, \emptyset\}$ независимы. Теперь запишем элементарный исход в виде произведения событий ... $G_1 \dots G_j \dots R_k \dots$, которое получается, если под каждым гербом записать событие G , а под каждой решкой — событие R , снабдив их соответствующими номерами. По

другому можно сказать, что $G_1 = \{\Gamma\ldots\}$, $G_2 = \{\ldots\Gamma\ldots\}$, $G_3 = \{\ldots\Gamma\ldots\}$ и т. д. ('три точки' означает здесь 'много точек') Ввиду независимости σ -алгебр \mathfrak{A}_i ,

$$\begin{aligned}\mathbb{P}(\ldots\Gamma\ldots\Gamma\ldots) &= \ldots\mathbb{P}(G_i)\ldots\mathbb{P}(R_j)\ldots\mathbb{P}(R_k)\ldots = \\ &= p^{\{\text{число гербов}\}}(1-p)^{\{\text{число решек}\}}.\end{aligned}$$

Вероятности всех наборов (то есть элементарных исходов), в которых число гербов равно k , оказываются равными. Поэтому вероятность события {в n экспериментах выпало k гербов} равно числу входящих в него элементарных исходов, умноженному на вероятность одного исхода, то есть

$$\mathbb{P}\{\text{в }n\text{ экспериментах выпало }k\text{ гербов}\} = C_n^k p^k (1-p)^{n-k}.$$

Пример. 6 раз бросался на стол игральный кубик (со сторонами 1, 2, 3, 4, 5, 6). Найти вероятность того, что 2 раза выпало число 1.

Решение. В данной задаче $p = 1/6$. Ответ: $C_6^2 (1/6)^2 (5/6)^4$.

18. Полиномиальная модель, в которой вместо двух исходов в биномиальной модели имеется r состояний, вероятность i -го состояния равна p_i , точно так же обобщает модель Больцмана — Максвелла. Соответствующие формулы можно вывести самостоятельно. В примере с кроликами мы можем считать разными размеры клеток, тогда вероятности попадания кролика в разные клетки будут разными и будут равняться соответственно p_1, \dots, p_r , каждый последующий кролик, занимая место в одной из клеток, не знает число кроликов, уже находящихся в клетке, поэтому события, связанные с разными кроликами, независимы в совокупности. В этой модели при рассадке $n = n_1 + \dots + n_r$ вероятность того, что для каждого i в i -ю клетку попадет ровно n_i кроликов, равна

$$\frac{n!}{n_1! \cdots n_r!} p_1^{n_1} \cdots p_r^{n_r}.$$

19. Пределная теорема Муавра — Лапласа (без доказательства).

Локальная предельная теорема Муавра — Лапласа имеет следующий вид:

$$\mathbb{P}\{S_n = k\} = C_n^k p^k (1-p)^{n-k} = \frac{1}{\sqrt{2\pi np(1-p)}} e^{-x(k)^2/2} (1 + \alpha_n),$$

где S_n — число гербов (т.е. сумма чисел 1 или 0) в n испытаниях в модели Бернулли, $x(k) = \frac{k - np}{\sqrt{np(1-p)}}$ и $\alpha_n \rightarrow 0$ равномерно при $n \rightarrow \infty$ для всех k , при которых $x(k)$ принадлежит какому-нибудь интервалу.

Интегральная предельная теорема Муавра — Лапласа имеет следующий вид:

$$\sum_{x(k) \in (a, b)} \mathbb{P}\{S_n = k\} = \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) (1 + \alpha_n) dx,$$

где $\alpha_n \rightarrow 0$. Здесь k имеет тот же порядок, что и n , следовательно, стремится к ∞ .

Мы не будем доказывать эти две теоремы (см. Гнеденко). Доказательство использует знаменитую формулу Стирлинга, которая будет нам полезна и в дальнейшем при оценке некоторых малых вероятностей:

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\theta(n)}, \text{ где } |\theta(n)| \leq \frac{1}{2n}$$

Локальная предельная теорема используется для приближенных вычислений очень малых вероятностей событий, связанных с моделью Бернулли, а интегральная теорема используется для приближенных вычислений уже немалых сумм вероятностей в той же схеме. В обоих случаях мы предполагаем при применении, что параметр p не слишком близок к 0 или 1,

а параметры n и k достаточно велики (однако неправильно организованные вычисления на компьютере этих приближений могут привести к машинному нулю). В дальнейшем, используя технику характеристических функций (преобразований Фурье распределений), будут доказаны результаты 20 века, очень частным случаем которых является интегральная теорема Муавра – Лапласа. Разумеется, стремительное развитие компьютерной техники уменьшает значимость подобных применений. Для p , близких к 0, и фиксированного не очень большого k хорошее приближение дает:

20. Предельная теорема Пуассона.

Пусть при $n \rightarrow \infty$ имеем $p = \frac{\lambda}{n} + o\left(\frac{1}{n}\right)$. Тогда для любого натурального k имеем:

$$\mathbb{P}\{S_n = k\} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}.$$

Доказательство.

$$\begin{aligned} C_n^k p^k (1-p)^{n-k} &= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n} + o\left(\frac{1}{n}\right) \right)^k \\ &\quad \left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right) \right)^{n-k} = \\ &1 \cdot \left(1 - \frac{1}{n} \right) \cdots \left(1 - \frac{k-1}{n} \right) \cdot \frac{1}{k!} \cdot (\lambda + o(1))^k \left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right) \right)^{n-k} \rightarrow \\ &1 \cdot \frac{1}{k!} \cdot \lambda^k \cdot e^{-\lambda}. \end{aligned}$$

Для примера с электрической цепью нам было нужно следующее предложение, которые я забыл предварительно сформулировать.

21. Предложение. *Если события A_1, A_2, \dots, A_n независимы в совокупности, множества индексов $I \subset \overline{1, n}$ и $J \subset \overline{1, n}$ не пересекаются, событие $A(I)$ выражается теоретико-множественными операциями через все события A_i , где $i \in I$, событие $A(J)$ выражается теоретико-множественными операциями через все события A_i , где $i \in J$, то события $A(I)$ и $A(J)$ независимы.*

Этот факт удобнее рассматривать, введя понятие независимых σ -алгебр.

В предложении нам нужно доказать независимость в совокупности σ -алгебры $\mathfrak{A}(I)$, порожденной всеми событиями A_i , $i \in I$, и σ -алгебры $\mathfrak{A}(J)$, порожденной всеми событиями A_i , $i \in J$. Рассмотрим для примера σ -алгебру $\mathfrak{A}(I)$. Она состоит из конечного числа событий, а следовательно, в ней имеются события, называемые *атомами* – минимальные события, не являющиеся невозможными. В качестве таких событий мы берем все события вида $\prod_{i \in I} B_i$, где каждое B_i равно или A_i или $\overline{A_i}$. Очевидно, что эти атомы попарно несовместны и каждое событие A_i является объединением $2^{|I|-1}$ числа атомов. Но тогда объединением атомов является и каждое событие из σ -алгебру $\mathfrak{A}(I)$. То же верно и для σ -алгебры $\mathfrak{A}(J)$. Очевидно, что любые два множества вида $\prod_{i \in I} B_i$ и $\prod_{i \in J} B_i$ независимы. Это следует, например, из импликации $c' \Rightarrow a$). Независимость любых событий $A(I) \in \mathfrak{A}(I)$ и $A(J) \in \mathfrak{A}(J)$ следует теперь из представимости этих событий как сумм атомов и многократного применения правила 2) операций с независимыми событиями.

22. Гипотеза независимости. Пример. Пусть один из пяти данных билетов является выигрышным. Вы купили два билета из пяти и рассматриваете события

$$A_i = \{i\text{-й купленный вами билет является выигрышным}\}.$$

Будут ли события A_1, A_2 независимыми? Ответ очевиден и отрицателен, так как

$$\mathbb{P}(A_1) = \mathbb{P}(A_2) = \frac{1}{5}, \mathbb{P}(A_1 A_2) = 0 \neq \mathbb{P}(A_1) \mathbb{P}(A_2).$$

Теперь немного усложним ситуацию, имеется не 5, а 1000 билетов, из которых 200 являются выигрышными. В этом случае вероятности событий A_i остаются прежними, но

$$\mathbb{P}(A_1 A_2) = \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) = \frac{200}{1000} \frac{199}{999}.$$

Мы можем сказать, что события A_1 и A_2 почти независимы. Однако ситуация упрощается, если мы не знаем общее число билетов и оно довольно велико, но мы знаем, что пятая часть билетов — выигрышная. В этой ситуации у нас нет другого варианта, кроме принятия гипотезы независимости A_1 и A_2 . Используя гипотезу независимости, мы вычисляем:

$$\mathbb{P}\{\text{оба билета выигрышные}\} = \mathbb{P}(A_1 A_2) = \frac{1}{5} \frac{1}{5},$$

$$\mathbb{P}\{\text{ровно один билет выигрышный}\} = \mathbb{P}(A_1 \bar{A}_2 + \bar{A}_1 A_2) = 2 \cdot \frac{1}{5} \frac{4}{5},$$

$$\mathbb{P}\{\text{оба билета проигрышные}\} = \mathbb{P}(\bar{A}_1 \bar{A}_2) = \frac{4}{5} \frac{4}{5}.$$

Упражнение 6. Любая ли σ -алгебра является монотонным классом?

Упражнение 7. Игральный кубик бросается на стол. Будут ли независимы следующие события: {результат бросания делится на 2}, {результат бросания делится на 3}?

Упражнение 8. В условиях задачи 2 §2 проверьте — какие пары событий будут зависимы, а какие — независимы: $\{x < 1\}$, $\{y < 1\}$; $\{x < y\}$, $\{x + y < 1\}$; $\{x < y\}$, $\{x + y < 2\}$; $\{x + y < 1\}$, $\{x + y < 2\}$?

Упражнение 9. В σ -алгебре событий заданы три события, A , B , C , независимые в совокупности, вероятности которых отличны от 0 и 1. Из скольких элементов состоит алгебра, порожденная A , B , C ? Сравните с упражнением 1b.

Часть 2. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

§5. Случайные величины

Определение 1. Борелевской σ -алгеброй на числовой прямой называется наименьшая σ -алгебра $\mathfrak{B}(\mathbf{R})$, содержащая все интервалы вида $(-\infty, x)$, где $x \in \mathbf{R}$. Элементы борелевской σ -алгебры называются обычно *борелевскими множествами*. Борелевской функцией на числовой прямой \mathbf{R} называется такое отображение $f : \mathbf{R} \rightarrow \mathbf{R}$, что прообраз $f^{-1}(B)$ любого борелевского множества B является борелевским множеством.

Определение 2. Случайной величиной называется измеримое отображение $\xi : (\Omega, \mathfrak{A}, \mathbb{P}) \rightarrow (\mathbf{R}, \mathfrak{B}(\mathbf{R}))$, т.е. для $\forall B \in \mathfrak{B}(\mathbf{R})$ прообраз $\xi^{-1}(B) = \{\omega \in \Omega : \xi(\omega) \in B\} \in \mathfrak{A}$, т.е. является событием.

Для сведения. Мы используем обозначения:

$\{\xi \in B\} := \{\omega : \xi(\omega) \in B\}$, $\{\xi < a\} := \{\omega : \xi(\omega) < a\}$, в том же духе понимаются обозначения $\{\xi > a\}$, $\{\xi = a\}$, $\{\xi \leq a\}$, $\{a \leq \xi < b\}$ и т.д.

В качестве *примера* применения таких обозначений удобнее рассмотреть (см. следующий параграф) случай нескольких случайных величин. Рассмотрим, например, эксперимент с тремя подбрасываниями монеты. Пространство элементарные исходов Ω состоит из 8 элементарных исходов, каждый из которых — набор из трех гербов или решек. Оцифруем этот эксперимент, для этого введем случайные величины ξ_i правилом:

$$\xi_i(\omega) = \begin{cases} 1, & \text{если в наборе } \omega \text{ на } i\text{-м месте герб,} \\ 0, & \text{если в наборе } \omega \text{ на } i\text{-м месте решка.} \end{cases}$$

Тогда, например, $\{(\xi_1, \xi_2, \xi_3) = (1, 1, 0)\} = \Gamma\Gamma R$, а

$$\{\xi_1 + \xi_2 + \xi_3 = 2\} = \{\Gamma\Gamma R, \Gamma R G, R\Gamma G\}.$$

Пример. Очень удобный примером в качестве вероятностного пространства является отрезок $[0, 1]$, в котором события — борелевские подмножества $[0, 1]$, а вероятность — сужение на отрезок меры Лебега λ . Случайными величинами оказываются тогда любые измеримые по Борелю функции из $[0, 1]$ в \mathbf{R} .

Замечания и упражнения. Проверьте, что всегда

$$\{\xi \in [0, 1/2]\} + \{\xi \in (1/2, 1]\} = \{\xi \in [0, 1]\}.$$

Писать $\{\xi \in [0, 1/2]\} + \{\xi \in [1/2, 1]\}$ мы уже не имеем права, так как может быть, что

$$\{\xi \in [0, 1/2]\} \cap \{\xi \in [1/2, 1]\} \neq \emptyset.$$

Это зависит от случайной величины ξ . Например, если $\Omega = \{\xi \in \{1, 2\}\}$, то

$$\{\xi \in [0, 1/2]\} \cap \{\xi \in [1/2, 1]\} = \emptyset.$$

Но всегда корректна формула

$$\{\xi \in [0, 1/2]\} \cup \{\xi \in [1/2, 1]\} = \{\xi \in [0, 1]\}.$$

Проверьте, что

$$\bigcup_n \{\xi \leq -1/n\} = \{\xi < 0\}.$$

Как проще записать $\prod_n \{-1/n < \xi < 1/n\}$?

Другое *определение 2'* случайной величины — такая функция $\xi : \Omega \rightarrow \mathbf{R}$, что событиями являются все множества вида $\{\xi < a\}$, $a \in \mathbf{R}$.

Предложение 3. *Определение 2 \Leftrightarrow Определение 2'.*

$1 \Rightarrow 1'$ очевидно так как $\{\xi < a\} = \xi^{-1}(-\infty, a)$, а $(-\infty, a)$ – это борелевское множество.

$1' \Rightarrow 1$. Рассмотрим класс \mathfrak{B}' таких множеств B в \mathbf{R} , что $\{\xi \in B\}$ является событием и докажем, что класс \mathfrak{B}' является σ -алгеброй. Например, покажем, что класс \mathfrak{B}' замкнут относительно счетного объединения. Пусть $B_n \in \mathfrak{B}'$, тогда

$$\left\{ \xi \in \bigcup_n B_n \right\} = \bigcup_n \{\xi \in B_n\} \in \mathfrak{A},$$

это означает, что $\bigcup_n B_n \in \mathfrak{B}'$.

Итак, \mathfrak{B}' – σ -алгебра. Кроме того, по построению \mathfrak{B}' содержит все интервалы вида $(-\infty, a)$. По определению борелевской σ -алгебры как наименьшей с этим же свойством \mathfrak{B}' содержит борелевскую σ -алгебру.

Упражнение. В предложении 3 вместо класса всех интервалов вида $(-\infty, x)$ взять любой другой класс множеств, порождающий борелевскую σ -алгебру. Сформулируйте и докажите это обобщение предложения 3.

Определение 4. Распределением \mathbb{P}_ξ случайной величины ξ называется вероятностная мера на борелевской σ -алгебре $\mathfrak{B}(\mathbf{R})$, которая задается соотношением

$$\mathbb{P}_\xi(B) := \mathbb{P}\{\xi \in B\} \text{ для всех } B \in \mathfrak{B}(\mathbf{R}). \quad (1)$$

Замечание. Корректность определения 4 следует из определения 2, в соответствии с которым правая часть (1) всегда имеет смысл. На первый взгляд кажется, что $\mathbb{P}_\xi(B)$ и $\mathbb{P}\{\xi \in B\}$ – это одно и то же, только вводится другое обозначение. Это конечно не так, распределение – вероятность, заданная не на событиях, а на борелевских множествах. В следующем параграфе мы предложим пример с фабрикой и мастерской, иллюстрирующий смысл перехода от случайной величины и вероятности на событиях к распределению случайной величины.

Аксиомы вероятности для \mathbb{P}_ξ легко выводятся из этих же аксиом для \mathbb{P} . Например, счетная аддитивность \mathbb{P}_ξ доказывается с помощью непосредственно проверяемого равенства событий

$$\left\{ \xi \in \sum_n B_n \right\} = \sum_n \{\xi \in B_n\},$$

где все B_n (дополнительная нагрузка на символ) попарно не пересекаются, и отсюда следует, что события $\sum_n \{\xi \in B_n\}$ также попарно не пересекаются.

Определение 5. Функцией распределения случайной величины ξ называется функция $F_\xi : \mathbf{R} \rightarrow [0, 1]$, задаваемая соотношением

$$F_\xi(x) = \mathbb{P}\{\xi < x\} \text{ для всех } x \in \mathbf{R}. \quad (2)$$

Каждая вероятность \mathbb{P} на числовой прямой задает функцию распределения $F(x) = \mathbb{P}(-\infty, x) \geq 0$, которая имеет следующие свойства:

- 1) $x \leq y$ влечет $F(x) \leq F(y)$;
- 2) $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$;
- 3) $\lim_{x \uparrow y} F(x) = F(y)$ для всех $y \in \mathbf{R}$.

Свойство 1) очевидно и следует из $(-\infty, x) \subset (-\infty, y)$.

Применим предложение 4.11 к доказательству свойств 2) и 3) функции распределения.

- 2) Пусть $x_n \downarrow -\infty$, тогда $(-\infty, x_n) \downarrow \emptyset$ и тогда

$$F(x_n) = \mathbb{P}(-\infty, x_n) \downarrow \mathbb{P}(\emptyset) = 0.$$

Точно так же рассматривается случай $x_n \uparrow +\infty$ и 3), т.е. рассматривается последовательность $x_n \uparrow x$, в этом случае $(-\infty, x_n) \uparrow (-\infty, x)$. Заметим, что в случае $x_n \downarrow x$ имеет место сходимость $(-\infty, x_n) \downarrow (-\infty, x]$, поэтому функция распределения не обязательно непрерывна справа.

Утверждение 6 (см. курс теории меры). *Если дана функция F , удовлетворяющая условиям 1), 2), 3) то существует единственная вероятность \mathbb{P} на борелевской σ -алгебре \mathfrak{B} такая, что $F(x) = \mathbb{P}(-\infty, x)$ для всех x .*

Для полноты изложения мы докажем эту теорему, но при этом будем ссыльаться на теорему о продолжении меры из функционального анализа, причем доказательство некоторых несложных фактов мы оставим читателю. Напомним, что начинается определение меры в курсе функционального анализа с определения меры на полуалгебре. Дополнительно, мы другим способом строго докажем единственность вероятности \mathbb{P} , заодно выясним, как по функции распределения F вычисляются вероятности некоторых важнейших числовых множеств на числовой прямой. Развитая при этом техника нам понадобится в дальнейшем при изучении понятия независимости случайных величин.

Итак, пусть дана полуалгебра \mathfrak{C} — это класс подмножеств некоторого большого множества Ω (которое часто называется *все пространство*) со следующими свойствами:

- 1) $\Omega \in \mathfrak{C}, \emptyset \in \mathfrak{C}$;
- 2) пересечение двух элементов \mathfrak{C} принадлежит \mathfrak{C} ;
- 3) если $A, B \in \mathfrak{C}, A \subset B$, то $B \setminus A$ представляется в виде объединения конечного числа попарно непересекающихся между собой элементов \mathfrak{C} (напомним, что если записать это свойство в виде формулы, то для объединения мы можем использовать знак суммы).

Напомним, что полуалгебра \mathfrak{C} порождает алгебру множеств \mathfrak{A} , причем каждое множество $A \in \mathfrak{A}$ представимо в виде $A = \sum_{i=1}^n A_i$, где $A_i \in \mathfrak{A} \forall i$.

Наконец, алгебра порождает σ -алгебру \mathfrak{B} .

Напомним также, что в теории меры предполагается заданной на полуалгебре \mathfrak{C} функция множеств $\mu : \mathfrak{C} \rightarrow [0, 1]$ со следующими свойствами: если

- 1) $\mu(\Omega) = 1$;
- 2) Если $A = \sum_n A_n$ (число слагаемых в сумме может быть как конечным, так и счетным), $A \in \mathfrak{C}, A_n \in \mathfrak{C} \forall n$, то

$$\mu(A) = \sum_n \mu(A_n).$$

Тогда согласно курсу функционального анализа функция множеств μ продолжается сначала на алгебру \mathfrak{C} , а потом на σ -алгебру \mathfrak{B} , каждый раз с сохранением свойства счетной аддитивности 3), причем в случае σ -алгебры условие $A \in \mathfrak{B}$ является уже излишним. Продолжение на \mathfrak{B} называется в нашем курсе вероятностью.

В нашем случае

$$\mathfrak{C} = \{[a, b) : a \in \mathbf{R} \cup \{-\infty\}, b \in \mathbf{R} \cup \{+\infty\}\},$$

функция μ на \mathfrak{C} определяется соотношением

$$\mu([a, b)) := F(b) - F(a).$$

Проверка свойства неотрицательности и 1) очевидна (поймите, что это действительно очевидно). А проверка свойства 2) очевидна лишь в случае конечного числа слагаемых в представлении A .

Ясно, что вероятности не всех числовых множеств нам интересны с практической точки зрения. Например, вряд ли кого-нибудь интересует вероятность множества всех рациональных чисел или вероятность канторова совершенного множества. В то же время вероятности разного рода интервалов и различных точек (которые являются замкнутыми интервалами) безусловно представляют интерес. Итак,

$$\mathbb{P}(-\infty, x) = F(x),$$

поэтому в силу конечной аддитивности \mathbb{P} должно быть верно

$$\mathbb{P}[a, b) = \mathbb{P}(-\infty, b) - \mathbb{P}(-\infty, a) = F(b) - F(a)$$

и эта вероятность однозначно определяется по F . Точно так же в силу конечной аддитивности для непересекающихся интервалов верно

$$\begin{aligned}\mathbb{P}([a_1, b_1) + [a_2, b_2) + \dots + [a_n, b_n)) &= \mathbb{P}[a_1, b_1) + \mathbb{P}[a_2, b_2) + \dots + \mathbb{P}[a_n, b_n) = \\ &= (F(b_1) - F(a_1)) + (F(b_2) - F(a_2)) + \dots + (F(b_n) - F(a_n)),\end{aligned}$$

т.е. на таких множествах \mathbb{P} также однозначно задается по F . Нетрудно проверить, что такие множества составляют алгебру множеств \mathfrak{A} (проверьте!).

Однозначность меры продолжается не только при сложении непересекающихся множеств или вычитании из большего множества меньшего, но и при переходе к пределу возрастающей или убывающей последовательности. По теореме 4.16 с помощью этих операций из элементов алгебры можно получить все элементы порожденной алгеброй σ -алгебры. В нашем случае это означает, что по функции распределения F вероятность на борелевской σ -алгебре восстанавливается однозначно.

7. Примеры. $\mathbb{P}(a, b) = F(b) - F(a+0)$; $\mathbb{P}[a, b] = F(b+0) - F(a)$; $\mathbb{P}\{x\} = F(x+0) - F(x)$ (т.е. вероятность точки равна скачку функции распределения в этой точке; если функция распределения непрерывна в точке, то вероятность этой точки равна нулю).

Предложение 8. Борелевская функция $f(\xi)$ от случайной величины ξ является случайной величиной.

Доказательство. $\{f(\xi) \in B\} = \{\xi \in f^{-1}(B)\}$. Для любого борелевского множества B множество $f^{-1}(B)$ также является борелевским множеством, по определению случайной величины $\{\xi \in f^{-1}(B)\}$ является событием.

Существенное замечание. Разумеется, вместо функции распределения, мы можем ввести другие функции, которые будут выполнять ту же роль экономного и видного на графике задания распределения вероятностей на числовой прямой, например,

$$F_\xi^*(x) = \mathbb{P}\{\xi \leq x\} \text{ для всех } x \in \mathbf{R}. \quad (2^*)$$

(эта функция непрерывна справа) или

$$G_\xi(x) = \mathbb{P}\{\xi > x\} \text{ для всех } x \in \mathbf{R}. \quad (2^{**})$$

(эта функция не возрастает и стремится к 1 на $-\infty$ и к 0 на $+\infty$).

9. Определение. Вероятностное распределение \mathbb{P} на борелевской σ -алгебре называется *дискретным*, если существует такое конечное или счетное множество чисел $\{x_n\}$, что $\sum_n \mathbb{P}\{x_n\} = 1$. В этом случае функция распределения имеет скачки в каждой точке x_n , равные $\mathbb{P}\{x_n\}$. Разумеется, сумма скачков равна 1.

10. Определение. Вероятностное распределение \mathbb{P} на борелевской σ -алгебре называется *распределением непрерывного типа*, если существует такая измеримая относительно борелевской σ -алгебры неотрицательная функция p на \mathbf{R} , что $p(x)dx = 1$ и функция p связана с вероятностью \mathbb{P} следующим соотношением:

$$\int_B p(x)dx = \mathbb{P}(B) \quad (3)$$

для любого борелевского множества B . Функция p называется плотностью распределения \mathbb{P} . Функция распределения F , определяющая распределение \mathbb{P} , в этом случае задается соотношением

$$\int_{-\infty}^x p(u)du = F(x) \quad (4)$$

(действительно, $B = (-\infty, x)$, интеграл Лебега по множеству меры нуль равен нулю, поэтому вероятность точки, заданная равенством (3), равна нулю, и нам все равно, интегрировать по интервалу $(-\infty, x)$ или по $(-\infty, x]$, мы вправе просто писать интеграл от $-\infty$ до x).

Обозначение. Если ξ — случайная величина, распределение которой имеет непрерывный тип, то плотность распределения обозначается p_ξ .

Алгоритм вычисления вероятности $\mathbb{P}(B)$ для дискретного и непрерывного типа распределения.

В дискретном случае распределение задается равенствами $\mathbb{P}\{a_n\} = p_n$, где $\sum_n p_n = 1$, $p_n \geq 0$ для всех n . Поэтому $\mathbb{P}\{a_1, a_2, \dots\} = 1$, $\mathbb{P}(\mathbf{R} \setminus \{a_1, a_2, \dots\}) = 0$. Поэтому вероятность пересечения любого борелевского множества B и $\mathbf{R} \setminus \{a_1, a_2, \dots\}$ равна 0. Чтобы вычислить $\mathbb{P}(B)$, надо пересечь его с $\{a_1, a_2, \dots\}$:

$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}(B \cap \{a_1, a_2, \dots\}) + \mathbb{P}[B \cap (\mathbf{R} \setminus \{a_1, a_2, \dots\})] = \\ &= \mathbb{P}(B \cap \{a_1, a_2, \dots\}) = \sum_{n: a_n \in B} p_n.\end{aligned}$$

В случае непрерывного типа распределения функция плотности обычно интегрируема по Риману и даже кусочно непрерывна, т.е. задается вида

$$p(x) = \sum_{i=1}^n p_i(x) I_{B_i}, \text{ где } \sum B_i = \mathbf{R}, \text{ все } B_i \text{ --- промежутки}.$$

В этом случае вероятность считается как сумма интегралов Римана:

$$\mathbb{P}(B) = \sum_i \int_{B_i \cap B} p_i(x) dx.$$

11. Примеры дискретных распределений

Распределение Бернулли задается соотношением

$$\mathbb{P}\{1\} = p, \mathbb{P}\{0\} = 1 - p.$$

Это распределение задается параметром $p \in (0, 1)$ и появляется в задаче о бросании монеты (если мы сопоставим гербу значение 1, а решке — 0 или наоборот).

Биномиальное распределение появляется в задаче о бросании монеты n раз и является распределением случайной величины — числа гербов в n испытаниях.

$$\mathbb{P}\{k\} = C_n^k p^k (1-p)^{n-k}, k = 0, 1, \dots, n.$$

Иногда для краткости используют символ распределения — $B(n, p)$, в котором кратко указаны вид распределения и его параметры. Например, $B(6, 1/3)$, вычислите вероятность $\{2\}$.

Распределение Пуассона появилось у нас в предельной теореме Пуассона и задается формулой

$$\mathbb{P}\{k\} = \frac{\lambda^k}{k!} e^{-\lambda}, k \in \{0\} \cup \mathbf{N}.$$

Символ этого распределения — $P(\lambda)$.

Гипергеометрическое распределение (см. задачу о выборе деталей).

$$\mathbb{P}\{r\} = \frac{C_R^r \cdot C_{N-R}^{n-r}}{C_N^n},$$

где $r \in [\max\{0, n - (N - R)\}, \min\{R, n\}] \cap (\{0\} \cup \mathbf{N})$.

Геометрическое распределение.

$$\mathbb{P}\{k\} = p(1-p)^k, k = 0, 1, 2, \dots, \text{ где } p \in (0, 1).$$

Геометрическое распределение появляется в следующей задаче: гнутая монета (вероятность выпадения герба равна p) подбрасывается до тех пор, пока не появится герб, $k + 1$ — число подбрасываний монеты.

12. Примеры распределений непрерывного типа

Равномерное распределение. Равномерное распределение на $[a,b]$ (неважно, отрезок или интервал) имеет функцию плотности

$$p(x) = \frac{1}{b-a} I_{[a,b]}(x),$$

где I_E — индикатор множества E — функция, равная 1 на E и 0 вне E .

Равномерное распределение появляется в геометрической модели. Параметры равномерного распределения — числа a, b или числа $\mu = \frac{a+b}{2}$, $\Delta = \frac{b-a}{2}$.

Экспоненциальное распределение. Плотность экспоненциального распределения задается формулой

$$p_\alpha(x) = \alpha e^{-\alpha x} I_{[0,\infty)}(x).$$

Иногда в качестве параметра используется $\beta = 1/\alpha$. Экспоненциальное распределение появилось в задаче о времени жизни атома.

Стандартное нормальное распределение задается плотностью

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Неплохо было бы убедиться в том, что интеграл от этой плотности по \mathbf{R} равен 1 и найти в Демидовиче интеграл Эйлера — Пуассона.

Нормальное распределение с произвольными параметрами. Если ξ имеет стандартное нормальное распределение, то случайная величина $\sigma\xi + m$ имеет нормальное распределение с параметрами m и $|\sigma|$. Надо научиться вычислять плотность распределения этой случайной величины по плотности ξ .

Покажем, как это можно сделать, используя связь плотности с функцией распределения. Пусть $\sigma > 0$.

$$\mathbb{P}\{\sigma\xi + m < x\} = \mathbb{P}\left\{\xi < \frac{x-m}{\sigma}\right\}.$$

Дифференцируем это равенство по формуле сложной функции. Получаем

$$p_{\sigma\xi+m}(x) = \frac{1}{\sigma} p_\xi\left(\frac{x-m}{\sigma}\right),$$

$$p_{\sigma\xi+m}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}.$$

Случай отрицательного σ аналогичен, но мы учитываем, что вероятность точки равна нулю. Дифференцируется равенство

$$\mathbb{P}\{\sigma\xi + m < x\} = \mathbb{P}\left\{\xi > \frac{x-m}{\sigma}\right\} = 1 - \mathbb{P}\left\{\xi < \frac{x-m}{\sigma}\right\}.$$

Получаем ту же плотность, что и для параметров $m, |\sigma|$.

$$p_{\sigma\xi+m}(x) = \frac{1}{-\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}.$$

Символ нормального распределения — $N(m, \sigma^2)$, стандартного нормального распределения — $N(0, 1)$. Нормальное распределение используется для моделирования многих ситуаций. Оно используется также для приближения других распределений.

Логарифмически-нормальное распределение. Это распределение случайной величины e^ξ , где ξ имеет нормальное распределение. Плотность этой величины выводится выше описанным способом:

$$\mathbb{P}\{e^\xi < x\} = \mathbb{P}\{\xi < \ln x\}, x > 0.$$

$$p_{e^\xi}(x) = \frac{1}{x} p_\xi(\ln x) I_{(0, \infty)}.$$

Распределение χ^2_1 . Аналогично выводится плотность случайной величины ξ^2 , где $\xi \sim N(0, 1)$. Но

$$\mathbb{P}\{-\sqrt{x} < \xi < \sqrt{x}\}, x > 0,$$

представляется как разность значений функции распределения ξ в двух точках — $-\sqrt{x}$ и \sqrt{x} .

Это распределение является частным случаем распределения χ^2_n при $n = 1$. В следующем параграфе мы рассмотрим распределение χ^2_n и некоторые другие распределения, связанные с наборами независимых нормальных случайных величин.

Распределение Коши.

Распределение Коши с параметрами a и μ имеет функцию плотности

$$p(x) = \frac{1}{\pi} \cdot \frac{a}{1 + a^2(x - \mu)^2}.$$

Распределение константы a . Мы забыли рассмотреть распределение константы a . Его часто называют мерой Дирака и обозначают $\delta_{\{a\}}$. Физики предпочитают использовать это обозначение для функции Дирака — обобщенной функции, которая равна везде, кроме a , нулю, в a равна $+\infty$, а интеграл от нее по числовой прямой равен 1. В этих терминах они дискретному распределению ($\mathbb{P}\{x_i\} = p_i$) приписывают плотность (обобщенную !) $p(x) = \sum_i p_i \delta_{\{x_i\}}(x)$. К распределению константы сходятся (определение сходимости будет существенно позже) многие непрерывные и дискретные распределения — Бернули (когда p сходится к 0 или 1), равномерное (когда b сходится к a), нормальное (когда σ сходится к 0) и др.

Наконец, кроме дискретных или непрерывных распределений возможны также *сингулярные распределения*. Вся вероятность сингулярного распределения сосредоточена на множестве лебеговой меры нуль, но вероятность каждой отдельной точки равна нулю. Трудно представить себе практическую ситуацию, где мы будем иметь дело с сингулярным распределением. Однако сейчас мы построим теоретически интересный пример сингулярного распределения. Это образ меры Лебега на отрезке $[0, 1]$ относительно отображения, переводящего любую точку $0 \dots 1 \dots 0 \dots$ этого отрезка в двоичной системе счисления в точку $0 \dots 2 \dots 0 \dots$ в троичной системе счисления. Отображение строится заменой каждой цифры 1 на цифру 2. Такое отображение переведет меру Лебега на канторово совершенное множество. Хорошо известно, что мера Лебега этого множества равна 0 (ее можно представить как $\lim_n (\frac{2}{3})^n$).

Симметричные распределения. Распределение случайной величины ξ называется симметричным, если распределение ξ совпадает с распределением случайной величины $-\xi$. В дискретном случае это означает, что

$$\mathbb{P}_\xi\{x_n\} = \mathbb{P}_\xi\{-x_n\} \text{ для всех значений } \xi,$$

а в непрерывном случае, что

$$p_\xi(x) = p_\xi(-x) \text{ для всех } x \in \mathbf{R}.$$

Легко видеть, что в случае симметричного распределения функция распределения F удовлетворяет равенству

$$F_\xi(x) = 1 - F_\xi(-x) \text{ для всех } x \in \mathbf{R}.$$

Поэтому таблицы функций симметричных распределений обычно содержат лишь значения $x > 0$. Примеры симметричных распределений — равномерное распределение на отрезке $[-a, a]$, нормальное распределение с $m = 0$, распределение Коши с $\mu = 0$.

Вопрос. Для какой константы распределение константы симметрично? Докажите сформулированные выше свойства функции плотности и функции распределения для симметричного распределения.

13. Как сконструировать случайную величину с заданным распределением \mathbb{P} , определенным функцией распределения F ?

Для этого мы должны определить некоторое вероятностное пространство $(\Omega, \mathcal{A}, \mathbb{P})$, а на нем определить случайную величину $\xi : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$, что $F_\xi = F$. Очевидно, что не всякое вероятностное пространство может подойти. Например, если Ω состоит из конечного или счетного множества точек, то распределение любой случайной величины на нем может быть лишь дискретным (Почему?).

Самый простой способ решения задачи тривиален — мы просто берем в качестве вероятностного пространства $(\mathbf{R}, \mathcal{B}(\mathbf{R}), \mathbb{P})$ и определяем ξ соотношением $\xi(x) = x$, \mathbf{R} . Другой способ позволяет использовать более стандартное вероятностное пространство — в качестве Ω мы берем интервал $(0, 1)$, в качестве σ -алгебры событий — σ -алгебру борелевских множеств в $(0, 1)$, а в качестве \mathbb{P} мы берем сужение меры Лебега m на числовой прямой на борелевские множества $(0, 1)$. Напомним, что мера m задается равенством $m(a, b) = b - a$ для всех чисел $a < b$. А теперь придумаем ξ . Дополним график функции распределения F (если эта функция имеет разрывы), включив разрывы в график. Таким образом, если $F(x) = a$, $F(x + 0) = b > a$, то в график включаются все точки (x, ω) , где $a < \omega \leq b$. Далее через каждую точку $\omega \in (0, 1)$ на оси y -в мы проводим прямую, параллельную оси x -ов плоть до пересечения $x = x(\omega)$ с нашим расширенным графиком (если таких точек пересечения бесконечно много, мы возьмем среди них самую левую точку). Итак, $\xi(\omega) := x(\omega)$.

Упражнение. Покажите, что в пункте 13 действительно $F_\xi = F$.

14. Таблицы вероятностей.

Для вычисления распределений служат таблицы вероятностей (напр. [30]). В них приводятся значения либо функции распределения $F(x)$ для большого числа x , функции $1 - F(x)$. Для эффективного использования таблиц, нужно уметь сводить распределения данного вида, но с произвольными параметрами, к табличным распределениям. Например, имеются таблицы для нормального распределения со средним 0 и дисперсией 1. Так как функция плотности p этого распределения симметрична,

$$\int_{-\infty}^{-a} p(x)dx = \int_a^{\infty} p(x)dx \text{ для любого } a,$$

т.е. функция распределения Φ в этом случае удовлетворяет соотношению $\Phi(x) = 1 - \Phi(-x)$. Поэтому многие таблицы функции Φ ограничиваются значениями $x > 0$. Функция распределения нормальной случайной величины ξ со средним m и дисперсией σ^2 легко вычисляется по таблицам Φ , так как случайная величина $\frac{\xi - m}{\sigma}$ имеет функцию распределения Φ .

Упражнение 10. Дано, что $\mathbb{P}\{\xi = 1\} = 1/3$, $\mathbb{P}\{\xi = 2\} = 1/2$, $\mathbb{P}\{\xi = 3\} = 1/6$. Найти $\mathbb{P}\{\xi \in \mathbf{R} \setminus \{1, 2, 3\}\}$ и построить график функции распределения F_ξ .

Упражнения 11. а. Дано, что $\Phi(x) - \Phi(-x) = 0.9$. Найти $\Phi(x)$.

б. Как по таблице значений функции распределения стандартной нормальной случайной величины ξ для всех $x > 0$ найти вероятности: $\mathbb{P}\{\xi < -2\}$, $\mathbb{P}\{|\xi| > 2\}$?

с. Найдите такое x , что $\mathbb{P}\{|\xi| > x\} = 0.1$.

Упражнение 12. Случайная величина ξ нормальна с $m = 1$ и $\sigma^2 = 4$. Найти по таблицам Φ величину $\mathbb{P}\{\xi \in (1, 2.5)\}$.

Задача. Как придумать случайную величину ξ такую, что $F_\xi = F$, где F имеет свойства 1)-3)? В качестве вероятностного пространства мы берем отрезок $[0, 1]$ с мерой Лебега λ , $\xi(\omega) =$

$F^{-1}(\omega)$. Заметим, что если $F(x) = \omega$ для x , пробегающих интервал, $F^{-1}(\omega)$ не существует, но множество таких ω не более чем счетно. Поэтому мера λ таких ω равна 0. Также, если функция распределения проскакивает x , то $\xi(\omega) = x$ для всех ω в отрезке $[F(x), F(x+0)]$. Пример другой случайной величины η с той же функцией распределения — $\eta(\omega) := \xi(1 - \omega)$.

Вопрос. Придумайте пример разных случайных величин с одинаковым распределением.

Намек. Прежде чем придумывать случайные величины, придумайте вероятностное пространство.

§6. Случайные векторы, наборы случайных величин

Рассматриваются случайные величины — числовые функции на пространстве всех элементарных исходов. Обозначения — ξ , η , ζ (мы будем использовать греческие буквы, в некоторых книгах используются большие латинские буквы — X , Y , ...). Простейший пример случайной величины — число гербов при n подбрасываниях монеты. В этом случае элементарные исходы — это всевозможные наборы ..Г..Р.. из гербов и решек длины n . Наша случайная величина (ее обычно обозначают S_n) сопоставляет каждому такому набору число гербов в наборе. Так как мы умеем считать вероятность каждого набора ($p^{S_n}(1-p)^{n-S_n}$), где p — вероятность выпадения герба в одном испытании, то мы можем подсчитать вероятность любого значения k случайной величины S_n , сложив вероятности всех наборов, на которых S_n принимает значение k . Как мы уже знаем, получается $C_n^k p^k (1-p)^{n-k}$. Более того, мы можем для любого множества B на числовой прямой найти вероятность $\mathbb{P}\{S_n \in B\}$ того, что S_n принадлежит B . Для этого нам нужно сложить вероятности всех k , которые содержатся в B . Заметим, что если мы незначительно изменим B , не включая в него новые точки k и не убирая их, то наша вероятность не изменится. Вероятность $\mathbb{P}\{S_n \in B\}$, заданная на всех B , называется распределением \mathbb{P}_{S_n} случайной величины S_n . Таким образом, мы меняем лишь обозначение:

$$\mathbb{P}_{S_n}(B) = \mathbb{P}\{S_n \in B\},$$

в частности,

$$\mathbb{P}_{S_n}(\{k\}) = \mathbb{P}\{S_n = k\},$$

но вводя новое обозначение распределения \mathbb{P}_{S_n} , мы забываем о случайных исходах — всех этих наборах из гербов и решек, и сохраняем информацию лишь о вероятностях всех множеств $\{S_n = k\}$. То же распределение, что и S_n , могут иметь и другие случайные величины, может быть даже заданные на других пространствах элементарных исходов. С распределением S_n связывается функция распределения

$$F_{S_n}(x) := \mathbb{P}_{S_n}(-\infty, x),$$

со значениями в отрезке $[0, 1]$. По этой функции легко восстанавливается само распределение. Нетрудно видеть, что вероятность $\mathbb{P}_{S_n}(-\infty, x)$ меняется лишь тогда, когда $(-\infty, x)$ захватывает очередное k , при этом увеличивается на вероятность этого k . Таким образом, зная величины скачков функции F_{S_n} , мы узнаем вероятности всех точек k , а по ним однозначно определяется распределение \mathbb{P}_{S_n} .

Возможны и более сложные ситуации, когда мы знаем, что некоторые величины случайны, но не можем построить пространство элементарных исходов, или же не можем его обозреть. Пусть, например, мы рассматриваем координату v_x скорости молекулы в газе. В рамках классической механики для нее используется следующая формула:

$$\mathbb{P}\{v_x \in (a, b)\} = \int_a^b \frac{\mu}{\sqrt{2\pi}} e^{-\frac{\mu^2 u^2}{2}} du, \quad (5)$$

где μ характеризует свойства газа. В этой формуле функция распределения случайной величины v_x имеет вид

$$F_{v_x}(s) := \int_{-\infty}^s \frac{\mu}{\sqrt{2\pi}} e^{-\frac{\mu^2 u^2}{2}} du, \quad (6)$$

(здесь мы не используем x , так как знак x указывает на направление координаты). Заметим, что эта формула противоречит конечности числа N молекул газа, функция распределения не растет скачками величины $1/N$. Такие распределения мы называем распределениями непрерывного типа (в отличие от распределений дискретного типа, у которых вероятность сосредоточена на конечном или счетном множестве точек). Подынтегральная функция в (5) или (6) называется функцией плотности, для случайной величины v_x она обозначается p_{v_x} .

Свойства функции плотности p_ξ случайной величины ξ очевидны. Во первых, $p_\xi \geq 0$ (хоть и бесконечно малая вероятность, но все же вероятность), во-вторых, $\int_{\mathbb{R}} p_\xi(x) dx = 1$. Однако в отличие от дискретного случая мы теперь уже не можем говорить о $\mathbb{P}\{v_x \in B\}$ для всех множеств B на числовой прямой, а только для таких B , для которых имеет смысл интеграл

$$\int_B \frac{\mu}{\sqrt{2\pi}} e^{-\frac{\mu^2 u^2}{2}} du = \mathbb{P}\{v_x \in B\}.$$

Однако для всех интересных для применений множеств B (обычно это интервалы) такой интеграл имеет смысл.

Замечание 1. Современные физики предпочитают все вероятностные распределения записывать как распределения непрерывного типа, используя в качестве плотности обобщенную функцию. Например, плотность дискретного распределения, заданного равенствами $\mathbb{P}\{a_i\} = p_i$, где $\sum_i p_i = 1$, можно определить обобщенной функцией $p(x) = \sum_i p_i \delta(x - a_i)$, где δ – функция Дирака.

Из определений следует, что в случае непрерывного типа распределения функция распределения F_ξ и функция плотности p_ξ связаны соотношениями:

$$F_\xi(x) = \int_{-\infty}^x p_\xi(t) dt,$$

отсюда дифференцированием выводится обратное соотношение

$$p_\xi(x) = \frac{dF_\xi}{dx}(x).$$

1. Многомерные распределения.

Разумеется, мы можем иметь дело не только с одной случайной величиной (в примерах выше S_n или v_x), но и одновременно с несколькими случайными величинами (например, при рассмотрении молекул газа мы одновременно иметь дело с тремя импульсами, p_x, p_y, p_z (механики обычно предпочитают скорости импульс как переменную) и тремя координатами q_x, q_y, q_z).

Итак, мы рассматриваем несколько случайных величин ξ_1, \dots, ξ_n и их совместное распределение $\mathbb{P}_{\xi_1, \dots, \xi_n}$, которое определяется соотношением:

$$\mathbb{P}_{\xi_1, \dots, \xi_n}(B) = \mathbb{P}\{(\xi_1, \dots, \xi_n) \in B\}.$$

Здесь B опять пробегает все полезные и хорошие множества, аналогами интервалов являются n -мерные прямоугольные параллелепипеды, но кроме них для приложений полезны и шары, пирамиды и другие множества.

Замечание. Как и в одомерном случае, мы используем обозначение

$$\{(\xi_1, \dots, \xi_n) \in B\} := \{\omega \in \Omega : (\xi_1(\omega), \dots, \xi_n(\omega)) \in B\}.$$

Заметим также, что

$$\{\xi_1 \in B_1, \dots, \xi_n \in B_n\} = \{\xi_1 \in B_1\} \dots \{\xi_n \in B_n\} \left(:= \bigcap_{i=1}^n \{\xi_i \in B_i\} \right).$$

Точнее, в качестве B мы можем использовать все элементы борелевской σ -алгебры – наименьшей σ -алгебры, содержащей шары со всеми центрами и всеми радиусами. Впрочем, многомерную борелевскую σ -алгебру мы можем также интерпретировать как наименьшую σ -алгебру, содержащую все множества вида

$$\{(x_1, \dots, x_n) : x_i < x\}, \text{ где } i = 1, 2, \dots, n, x \text{ пробегает } \mathbf{R},$$

то есть полупространства. Теоретико-множественными операциями мы можем получить из полупространств многомерные параллелепипеды, а из них счетными объединениями открытые шары.

Пример, иллюстрирующий целесообразность рассмотрения вместо случайной величины ее распределения. Директор модного ателье должен иметь список клиентов, у каждого из которых свои параметры одежды. Таким образом, директор должен иметь вероятностное пространство, элементарными исходами которого являются клиенты. Случайные величины на этом пространстве сопоставляют каждому клиенту его параметры (например, рост, размер, полнота). Наконец, если все клиенты одинаково интересны для мастерской и заказывают примерно одно и то же количество одежды, на множестве клиентов можно ввести классическую вероятность, в которой все клиенты равновероятны.

Директора фабрики по производству одежды интересует только совместное распределение вероятностей у его потребителей таких случайных величин, как рост, размер, полнота. Полного списка клиентов он не знает, так как он слишком велик. Обычно директор имеет информацию о вероятностях таких событий, как например

$$\{3 \text{ рост, } 46 \text{ размер, маленькая полнота}\}.$$

Но если одежды каких-то размеров в избытке, а других не хватает, это по-видимому объясняется тем, что он использовал распределения вероятностей, справедливые для всего населения, применительно к какой-то особой группе населения с другими распределениями вероятностей.

Итак, переход от случайной величины или вектора к распределению означает обезличивание элементарных исходов.

Многомерные распределения могут быть как дискретными, так и непрерывными (а если использовать язык обобщенных функций, то можно и все мыслить непрерывными).

В многомерном случае задается также совместная функция распределения следующим соотношением:

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \mathbb{P}\{\xi_1 < x_1, \dots, \xi_n < x_n\}.$$

Но эта функция не так интересна как в одномерном случае: в многомерном случае нет хорошего порядка для точек, да и график такой функции нарисовать затруднительно. Совместная функция распределения имеет те же (с учетом многомерности) свойства 1)-3), что и одномерная функция распределения, но этих свойств недостаточно, чтобы определять многомерную вероятность. Например, в двумерном случае нужно, чтобы $F_{\xi, \eta}$ имела дополнительное свойство:

$$\begin{aligned} \mathbb{P}\{a_1 \leq \xi < a_2, b_1 \leq \eta < b_2\} = \\ = F_{\xi, \eta}(a_2, b_2) - F_{\xi, \eta}(a_1, b_2) - F_{\xi, \eta}(a_2, b_1) + F_{\xi, \eta}(a_1, b_1) \geq 0 \end{aligned}$$

для всех $a_1 < a_2, b_1 < b_2$.

Как и в случае одной случайной величины, совместное распределение может определяться совместной функцией плотности

$$p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n),$$

которая связана с распределением следующим соотношением:

$$\mathbb{P}_{\xi_1, \dots, \xi_n}(B) = \int \dots \int_B p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) dx_1 \dots dx_n. \quad (7)$$

Можно отметить также связь совместной функции распределения и совместной функции плотности:

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} p_{\xi_1, \dots, \xi_n}(u_1, \dots, u_n) du_1 \dots du_n.$$

Обратно,

$$p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \frac{\partial^n F_{\xi_1, \dots, \xi_n}}{\partial x_1 \dots \partial x_n}(x_1, \dots, x_n).$$

Вопрос. Пусть нам дана совместная плотность случайных величин x_1, \dots, x_n , как найти совместную плотность части этого набора — $\xi_{n_1}, \dots, \xi_{n_k}$, $1 \leq n_i \leq n$? Ответ следует из определения совместной функции плотности. Пусть B — k -мерное борелевское множество из $\prod_{i \in \{n_1, \dots, n_k\}} \mathbf{R}_i$, где \mathbf{R}_i числовая прямая под номером i (мы имеем в виду, что \mathbf{R}^n представляется в виде $\prod_{i \leq n} \mathbf{R}_i$.)

Итак,

$$\begin{aligned} \int \dots \int_B p_{\xi_{n_1}, \dots, \xi_{n_k}}(x_{n_1}, \dots, x_{n_k}) dx_{n_1} \dots dx_{n_k} &= \mathbb{P}\{(\xi_{n_1}, \dots, \xi_{n_k}) \in B\} = \\ &= \mathbb{P}\{(\xi_1, \xi_2, \dots, \xi_n) \in B \times \prod_{i \notin \{n_1, \dots, n_k\}} \mathbf{R}_i\} = \\ &= \int \dots \int_{B \times \prod_{i \notin \{n_1, \dots, n_k\}} \mathbf{R}_i} p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) dx_1 \dots dx_n. \end{aligned}$$

Отсюда следует, что в качестве $p_{\xi_{n_1}, \dots, \xi_{n_k}}$ (плотность задается с точностью до множества меры нуль) мы можем взять

$$\begin{aligned} p_{\xi_{n_1}, \dots, \xi_{n_k}}(x_{n_1}, \dots, x_{n_k}) &= \\ &= \int \dots \int_{\prod_{i \notin \{n_1, \dots, n_k\}} \mathbf{R}_i} p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) \prod_{i \notin \{n_1, \dots, n_k\}} dx_i. \end{aligned}$$

В многих ситуациях нужно переходить от совместной функции плотности $p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)$ к совместной функции $p_{\eta_1, \dots, \eta_n}(y_1, \dots, y_n)$, где случайный вектор (η_1, \dots, η_n) получен в результате преобразования (обычно непрерывного или кусочно непрерывного) случайного вектора (ξ_1, \dots, ξ_n) . Поэтому нужно уметь как-то записывать борелевское множество \tilde{B} через B в равенстве

$$\mathbb{P}\{(\xi_1, \dots, \xi_n) \in B\} = \mathbb{P}\{(\eta_1, \dots, \eta_n) \in \tilde{B}\},$$

(чтобы всегда, для всех вариантов совместных распределений, были равны вероятности, нужно, чтобы были равны события) после чего записать это равенство как равенство интегралов от соответствующих функций плотности:

$$\int \dots \int_B p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) dx_1 \dots dx_n = \int \dots \int_{\tilde{B}} p_{\eta_1, \dots, \eta_n}(y_1, \dots, y_n) dy_1 \dots dy_n.$$

Далее, используя замену переменных под знаком интеграла, можно попытаться выразить $p_{\eta_1, \dots, \eta_n}(y_1, \dots, y_n)$ через $p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)$. Мы используем то соображение, что из равенства интегралов для всех множеств B (интеграл как функцию области интегрирования B часто

называют неопределенным интегралом) следует равенство (с точностью до множества лебеговой меры нуль) подинтегральных функций. Приведем пример такого перехода: пусть заданы случайные векторы (ξ, η) , (φ, ψ) , где $\varphi = \xi^2 \eta$, $\psi = \xi/\eta$. Тогда чтобы обеспечить равенство

$$\{(\xi, \eta) \in B\} = \{((\varphi, \psi)) \in \tilde{B}\},$$

множество \tilde{B} должно состоять из всех точек (u, v) таких, что $(\sqrt[3]{uv}, \sqrt[3]{u/v^2}) \in B$.

Замечание. Пусть ξ, η – случайные величины, заданные на вероятностном пространстве $(\Omega, \mathfrak{A}, \mathbb{P})$ и имеющие совместное распределение $\mathbb{P}_{\xi, \eta}$. Рассмотрим функции $x : (x, y) \rightarrow x$, $y : (x, y) \rightarrow y$. Эти функции являются случайными величинами, заданными на вероятностном пространстве $(\mathbf{R}^2, \mathfrak{B}(\mathbf{R}^2), \mathbb{P}_{\xi, \eta})$. Очевидно (?), что $\mathbb{P}_{x, y} = \mathbb{P}_{\xi, \eta}$.

2. Независимые случайные величины.

Случайные величины ξ_1, \dots, ξ_n называются независимыми, если для любых борелевских множеств B_1, B_2, \dots, B_n имеет место равенство

$$\mathbb{P}\{\xi_1 \in B_1, \dots, \xi_n \in B_n\} = \mathbb{P}\{\xi_1 \in B_1\} \dots \mathbb{P}\{\xi_n \in B_n\}. \quad (8)$$

Можно доказать, это довольно сложно (мы это сделаем ниже), что для независимости случайных величин $\xi_1, \xi_2, \dots, \xi_n$ достаточно равенств

$$\mathbb{P}\{\xi_1 < a_1, \dots, \xi_n < a_n\} = \mathbb{P}\{\xi_1 < a_1\} \dots \mathbb{P}\{\xi_n < a_n\} \quad (9)$$

для всех вещественных чисел a_1, a_2, \dots, a_n , которые следуют из (8) при $B_i = (-\infty, x_i)$.

Заметим, что как (8), так и (9) на деле означают независимость в совокупности входящих в (8) или (9) наборов событий. В (8) некоторые из событий B_i мы можем заменить на \mathbf{R} , из определения случайной величины следует, что $\{\xi_n \in \mathbf{R}\} = \Omega$. Таким образом, некоторые из сомножителей в правой части окажутся равными 1, а некоторые события в правой части можно будет не писать, и мы получим равенство (8) для части набора событий. То же верно и для (9).

Борелевские функции $f(\xi_i)$ независимых случайных величин ξ_i сами независимы. Действительно, вместо (8) нам нужно доказать

$$\begin{aligned} \mathbb{P}\{f_1(\xi_1) \in B_1, \dots, f_n(\xi_n) \in B_n\} = \\ \mathbb{P}\{f_1(\xi_1) \in B_1\} \dots \mathbb{P}\{f_n(\xi_n) \in B_n\}. \end{aligned} \quad (8')$$

Но $\{f_i(\xi_i) \in B_i\} = \{(\xi_i) \in f_i^{-1}(B_i)\}$. Таким образом, (3') – это (3) при $B_i = f_i^{-1}(B_i)$.

Рассмотрим случай, когда $\xi_1, \xi_2, \dots, \xi_n$ – независимые случайные величины с дискретным распределением, то есть для любого k мы имеем: $\mathbb{P}\{\xi_k = a_i\} = p_i^{(k)}$, где $i \leq r^{(k)}$, $\sum_{i=1}^{r^{(k)}} p_i^{(k)} = 1$. Тогда

$$\mathbb{P}\{\xi_1 = a_{i_1}, \dots, \xi_n = a_{i_n}\} = p_{i_1}^{(1)} p_{i_2}^{(2)} \dots p_{i_n}^{(n)}.$$

Аналогично рассматривается случай независимых случайных величин $\xi_1, \xi_2, \dots, \xi_n$ с непрерывным типом распределения. Легко показывается, что в этом случае имеется совместная функция плотности $p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)$, которая задается равенством

$$p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = p_{\xi_1}(x_1) \dots p_{\xi_n}(x_n). \quad (10)$$

Это легко выводится дифференцированием равенства (9) по всем переменным. Обратно, интегрированием равенства (10) по всем переменным мы получаем (8) или (9). Таким образом, если *совместная функция плотности случайных величин ξ_1, \dots, ξ_n и функции плотности отдельных случайных величин связаны соотношением (10), то случайные величины ξ_1, \dots, ξ_n независимы*. Почему-то вызывает затруднения важный для математической статистики случай независимых случайных величин ξ_1, \dots, ξ_n с одинаковым распределением непрерывного типа с функцией плотности p . Разумеется, совместная функция плотности не p^n , а

$$p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = p(x_1) \dots p(x_n).$$

Теорема 3. Следующие условия эквивалентны (каждое из них означает независимость случайных величин): для любых борелевских множеств B_1, \dots, B_n имеет место равенство

$$\mathbb{P}\{\xi_1 \in B_1, \dots, \xi_n \in B_n\} = \mathbb{P}\{\xi_1 \in B_1\} \dots \mathbb{P}\{\xi_n \in B_n\}, \quad (11)$$

для всех вещественных чисел a_1, \dots, a_n имеет место равенство:

$$\mathbb{P}\{\xi_1 < a_1, \dots, \xi_n < a_n\} = \mathbb{P}\{\xi_1 < a_1\} \dots \mathbb{P}\{\xi_n < a_n\}. \quad (12)$$

Доказательство. Разумеется, (8) влечет (9). Обратное нужно доказывать, так как (9) означает справедливость (8) лишь для интервалов вида $(-\infty, a)$.

Напомним, что из независимости в совокупности двух наборов событий: $A_1, A_2, \dots, A_n, \tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_n$ таких, что A_1, \tilde{A}_1 несовместны (т.е. $A_1 \tilde{A}_1 = \emptyset$, следует независимость в совокупности и набора $A_1 + \tilde{A}_1, A_2, \dots, A_n$, аналогично проверяется, что если условие несовместности заменить на условие $A_1 \supset \tilde{A}_1$, то независим в совокупности набор $A_1 \setminus \tilde{A}_1, A_2, \dots, A_n$. Разумеется, то же верно и для пар A_i, \tilde{A}_i с другими номерами.

Мы используем это простые утверждения для наборов событий вида $\{\xi_1 < b_1\}, \dots, \{\xi_n < a_n\}$ и $\{\xi_1 < a_1\}, \{\xi_2 < a_2\}, \dots, \{\xi_n < a_n\}$. Согласно (11) оба этих набора независимы в совокупности, а поэтому при $a_1 < b_1$ независимы в совокупности и события $\{a_1 < \xi_1 < b_1\}, \{\xi_2 < a_2\}, \dots, \{\xi_n < a_n\}$. Итак, из независимости прообразов относительно случайных величин ξ_1, \dots, ξ_n борелевских множеств вида $(-\infty, a)$ следует независимость прообразов множеств вида $[a_i, b_i]$. Аналогично, но уже используя операцию сложения, мы получаем независимость прообразов множеств вида

$$\sum_{j=1}^{n_i} [a_i^{(j)}, b_i^{(j)}]. \quad (13)$$

Такие множества составляют уже алгебру множеств в \mathbf{R} (если предполагать, что одно из чисел $a_i^{(j)}$ может равняться $-\infty$, а одно из чисел $b_i^{(j)}$ — ∞). Заметим, что нам удалось избежать операций пересечения и объединения множеств общего вида, при таких операциях (см. пример Бернштейна) независимость может потеряться.

Но нам нужно доказать независимость прообразов любых борелевских множеств в \mathbf{R} . Согласно теореме о монотонных классах для их получения из множеств вида (*) нам не нужно использовать операции счетного объединения или счетного пересечения, а достаточно использовать операции предела возрастающей последовательности множеств (здесь это объединение) или убывающей последовательности множеств (пересечение). Однако такие операции также сохраняют независимость. Это легко проверяется: если для любого m события $A_1^{(m)}, A_2, \dots, A_n$ удовлетворяют соотношению

$$\mathbb{P}(A_1^{(m)} A_2, \dots, A_n) = \mathbb{P}(A_1^{(m)}) \mathbb{P}(A_2) \dots \mathbb{P}(A_n), \quad (**)$$

а $A_1^{(m)} \nearrow A_1$ или $A_1^{(m)} \searrow A_1$, то переходом к пределу в (**) мы получаем

$$\mathbb{P}(A_1 A_2, \dots, A_n) = \mathbb{P}(A_1) \mathbb{P}(A_2) \dots \mathbb{P}(A_n).$$

Теорема 4. Пусть случайные величины (все вместе) $\xi_1, \xi_2, \dots, \xi_n, \eta_1, \eta_2, \dots, \eta_m$ независимы, $f(x_1, x_2, \dots, x_n)$ и $g(x_1, x_2, \dots, x_m)$ — борелевские функции. Тогда случайные величины $f(\xi_1, \dots, \xi_n), g(\eta_1, \dots, \eta_m)$ независимы.

Замечание. Для упрощения обозначений я рассматриваю набор независимых случайных величин, разделенный лишь на две пачки. Без изменений в доказательстве все это верно и для набора, разделенного на любого конечного числа пачек. К сожалению, нельзя ограничиться случаем $n = m = 1$, для него все очевидно и уже доказано.

Доказательство аналогично доказательству теоремы 1. Из условий теоремы немедленно следует независимость событий вида

$$\{(\xi_1, \xi_2, \dots, \xi_n) \in B_1 \times \dots \times B_n\}, \{(\eta_1, \eta_2, \dots, \eta_m) \in C_1 \times \dots \times C_m\},$$

где множества B_i , C_j являются борелевскими множествами. Входящие в эти множества произведения составляют полуалгебры множеств соответственно в \mathbf{R}^n и \mathbf{R}^m . Эти полуалгебры порождают алгебры множеств. Как мы уже видели, их прообразы относительно отображений $(\xi_1, \xi_2, \dots, \xi_n)$ и $(\eta_1, \eta_2, \dots, \eta_m)$ также независимы. Алгебры порождают борелевские σ -алгебры соответственно в \mathbf{R}^n и \mathbf{R}^m (проверьте для себя, что на плоскости любой открытый круг можно представить как объединение счетного числа открытых прямоугольников, открытые множества являются объединениями счетного числа шаров и т. д.). Таким образом, опираясь на теорему о монотонных классах, используя как и выше сохранение независимости при некоторых операциях с независимыми событиями, мы доказываем, что прообразы борелевских множеств относительно рассмотренных отображений являются независимыми событиями. Осталось отметить, что для любых борелевских множеств B и C множества $f^{-1}(B)$ и $g^{-1}(C)$ принадлежат борелевским σ -алгебрам в соответствующих пространствах.

Определение 5. Свертка функций распределения F_1 и F_2 — это функция распределения F суммы независимых случайных величин с функциями распределения F_1 и F_2 .

$$F(x) = \int_{-\infty}^{\infty} F_1(x-t)dF_2(t). \quad (14)$$

Доказательство. Это утверждение — следствие теоремы Фубини. Но в данном простом случае имеется простое доказательство, использующее теорему Лебега. По определению

$$F(x) = \iint_{u+t < x} dF_1(u)dF_2(t).$$

Мы представим множество $A = \{u+t < x\}$ как предел возрастающей последовательности сумм бесконечных тонких прямоугольников. Разобьем

$$\mathbf{R} = \sum_{k \in \mathbf{Z}} [k/2^n, k + 1/2^n),$$

тогда $A_n \uparrow A$, где

$$A_n = \sum_{k \in \mathbf{Z}} \{(t, u) \in [k/2^n, (k+1)/2^n) \times (-\infty, x - (k+1)/2^n)\}.$$

По построению

$$\begin{aligned} F^{(n)}(x) &:= \iint_{A_n} dF_1(u)dF_2(t) = \\ &= \sum_{k \in \mathbf{Z}} F_1(x - (k+1)/2^n)[F_2((k+1)/2^n) - F_2(k/2^n)] = \\ &= \int_{-\infty}^{\infty} F_1^{(n)}(x-t)dF_2(t), \end{aligned}$$

где $F_1^{(n)}(x-t) = F_1(x - (k+1)/2^n)$ для $t \in [k/2^n, k + 1/2^n)$. Очевидно, что $F_1^{(n)}(x-t) \rightarrow F_1(x-t)$ для всех t . Так как $A_n \uparrow A$, $F^{(n)}(x) \uparrow F(x)$, в то же время по теореме Лебега

$$\int_{-\infty}^{\infty} F_1^{(n)}(x-t)dF_2(t) \rightarrow \int_{-\infty}^{\infty} F_1(x-t)dF_2(t).$$

Сейчас мы введем очень важное для применений распределение — *стандартное нормальное n-мерное распределение*. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины со стандартным нормальным распределением, то есть с функцией плотности

$$p_{\xi_i}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Тогда

$$p_{\xi_1, \xi_2, \dots, \xi_n}(x_1, x_2, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{x_1^2 + x_2^2 + \dots + x_n^2}{2}}.$$

Очень важным для применений в математической статистике и любопытным является следующий факт:

6. Лемма о вращении стандартного нормального случайного вектора. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины со стандартным нормальным распределением, U — ортогональное преобразование (то есть вращение с центром в нуле) пространства \mathbf{R}^n . Тогда случайный вектор

$$(\eta_1, \eta_2, \dots, \eta_n) = U(\xi_1, \xi_2, \dots, \xi_n)$$

имеет ту же совместную функцию плотности, что и $(\xi_1, \xi_2, \dots, \xi_n)$, т. е. $\eta_1, \eta_2, \dots, \eta_n$ — независимые случайные величины со стандартным нормальным распределением.

Мы будем исходить из определения (*).

$$\begin{aligned} \mathbb{P}\{(\eta_1, \eta_2, \dots, \eta_n) \in B\} &= \mathbb{P}\{U(\xi_1, \xi_2, \dots, \xi_n) \in B\} = \\ &= \mathbb{P}\{(\xi_1, \xi_2, \dots, \xi_n) \in U^{-1}(B)\} = \\ &= \int \dots \int_{U^{-1}(B)} p_{\xi_1, \xi_2, \dots, \xi_n}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = \\ &= \int \dots \int_{U^{-1}(B)} \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{x_1^2 + x_2^2 + \dots + x_n^2}{2}} dx_1 dx_2 \dots dx_n. \end{aligned}$$

Произведем в этом интегrale замену переменных:

$$(y_1, y_2, \dots, y_n) = U(x_1, x_2, \dots, x_n).$$

Тогда область интегрирования $\{(x_1, x_2, \dots, x_n) \in U^{-1}(B)\}$ перейдет в область $\{(y_1, y_2, \dots, y_n) \in B\}$. Хорошо известно из алгебры, что якобиан при ортогональном вращении равен 1, а сумма квадратов координат переходит в сумму квадратов координат. Итак, мы имеем:

$$\begin{aligned} \mathbb{P}\{(\eta_1, \eta_2, \dots, \eta_n) \in B\} &= \\ &= \int \dots \int_B \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{y_1^2 + y_2^2 + \dots + y_n^2}{2}} dy_1 dy_2 \dots dy_n. \end{aligned}$$

Таким образом, совместная функция плотности осталась прежней.

Замечание. Как мы видим, в наших выкладках полезным приемом является переход от интеграла

$$= \int \dots \int_B p_{\xi_1, \xi_2, \dots, \xi_n}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

(с которым нам трудно работать) к вероятности события

$$\mathbb{P}\{(\xi_1, \xi_2, \dots, \xi_n) \in B\},$$

которое мы можем преобразовать, изменив вектор $(\xi_1, \xi_2, \dots, \xi_n)$ и борелевское множество B так, чтобы событие осталось прежним.

7. Следствие. Если ξ, η – независимые стандартные нормальные случайные величины, то случайные величины $\frac{\xi + \eta}{\sqrt{2}}, \frac{\xi - \eta}{\sqrt{2}}$ тоже независимы и имеют то же распределение.

Упражнение и Пример. Все это верно лишь для одинаково распределенных **нормальных** случайных величин. В качестве контрпримера рассмотрим независимые ξ, η с равномерным распределением на отрезке $[-1, 1]$. Вычислите и сравните вероятности событий $\{\xi + \eta \in [-1, +1]\}, \{\xi - \eta \in [-1, +1]\}, \{\xi + \eta \in [-1, +1], \xi - \eta \in [-1, +1]\}$

8. Невырожденное многомерное нормальное распределение.

Многомерное нормальное распределение можно определить как распределение случайного вектора $T(\xi_1, \xi_2, \dots, \xi_n)$, где $(\xi_1, \xi_2, \dots, \xi_n)$ – стандартный нормальный случайный вектор с совместной функцией плотности

$$p_{\xi_1, \xi_2, \dots, \xi_n}(x_1, x_2, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{x_1^2 + x_2^2 + \dots + x_n^2}{2}},$$

а T получен последовательным применением линейного преобразования S (справа!) и сдвига на вектор (m_1, \dots, m_n) . Таким образом, $T(\xi_1, \xi_2, \dots, \xi_n) = S(\xi_1, \xi_2, \dots, \xi_n) + (m_1, \dots, m_n)$. Вычислим функцию плотности случайного вектора $T(\xi_1, \xi_2, \dots, \xi_n)$, если она существует. Сначала освободимся от сдвига. Покажем, что

$$\begin{aligned} p_{\xi_1+m_1, \xi_2+m_2, \dots, \xi_n+m_n}(x_1, x_2, \dots, x_n) &= \\ &= p_{\xi_1, \xi_2, \dots, \xi_n}(x_1 - m_1, x_2 - m_2, \dots, x_n - m_n). \end{aligned}$$

Это следует из определения функции плотности. С одной стороны

$$\begin{aligned} \mathbb{P}\{(\xi_1 + m_1, \xi_2 + m_2, \dots, \xi_n + m_n) \in B\} &= \\ &= \int_B p_{\xi_1+m_1, \xi_2+m_2, \dots, \xi_n+m_n}(x_1, x_2, \dots, x_n) dx_1 \dots dx_n. \end{aligned}$$

С другой стороны

$$\begin{aligned} \mathbb{P}\{(\xi_1 + m_1, \xi_2 + m_2, \dots, \xi_n + m_n) \in B\} &= \\ &= \mathbb{P}\{(\xi_1, \xi_2, \dots, \xi_n) \in B - (m_1, \dots, m_n)\} = \\ &= \int_{B-(m_1, \dots, m_n)} p_{\xi_1, \xi_2, \dots, \xi_n}(x_1, x_2, \dots, x_n) dx_1 \dots dx_n. \end{aligned}$$

Далее будем считать, что сдвига уже нет, а преобразование S является невырожденным. Рассмотрим вероятность

$$\mathbb{P}\{(\xi_1, \xi_2, \dots, \xi_n)S \in B\} = \int_B p_{S(\xi_1, \xi_2, \dots, \xi_n)}(x_1, x_2, \dots, x_n) dx_1 \dots dx_n.$$

Имеем:

$$\begin{aligned} \mathbb{P}\{(\xi_1, \xi_2, \dots, \xi_n)S \in B\} &= \mathbb{P}\{(\xi_1, \xi_2, \dots, \xi_n) \in (B)S^{-1}\} = \\ &= \int_{(B)S^{-1}} \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{x_1^2 + x_2^2 + \dots + x_n^2}{2}} dx_1 \dots dx_n. \end{aligned}$$

Производим преобразование $(x_1, \dots, x_n) = (y_1, \dots, y_n)S$, нам также удобно использовать матричное обозначение для квадратичной формы – $x_1^2 + x_2^2 + \dots + x_n^2 = [x_1, \dots, x_n][x_1, \dots, x_n]'$. Имеем:

$$\int_{(B)S^{-1}} \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{x_1^2 + x_2^2 + \dots + x_n^2}{2}} dx_1 \dots dx_n =$$

$$\begin{aligned}
&= \int_{(B)S^{-1}} \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{[x_1, \dots, x_n][x_1, \dots, x_n]'}{2}} dx_1 \dots dx_n = \\
&= \int_B |\det(S)| \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{[y_1, \dots, y_n]SS'[y_1, \dots, y_n]'}{2}} dy_1 \dots dy_n = \\
&= \int_B \det(A)^{1/2} \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{\sum_{i,j} a_{ij} y_i y_j}{2}} dy_1 \dots dy_n,
\end{aligned}$$

где $A = SS' = [a_{ij}]_{i,j \leq n}$.

Очень важное замечание. Если матрица $[a_{ij}]$ диагональна, то все случайные величины ξ_i независимы.

Вопрос. $a_{ij} = 0$. Следует ли из этого независимость ξ_i и ξ_j ?

В дальнейшем мы встретим также вырожденное многомерное нормальное распределение, у которого нет функции плотности.

Интегрируя совместную функцию плотности по некоторым переменным (см. Вопрос) мы опять получаем функцию того же вида. Итак, часть набора случайных величин с невырожденным совместным нормальным распределением также имеет невырожденное совместное нормальное распределение.

Распределение χ^2_n .

Сейчас мы выведем распределение, очень важное для математической статистики. Пусть снова $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины со стандартным нормальным распределением, $\chi_n^2 = \xi_1^2 + \xi_2^2 + \dots + \xi_n^2$. Найдем плотность распределения χ^2 . Мы опять воспользуемся связью функции плотности и функции распределения и представим плотность как производную функции $\mathbb{P}\{\chi_n^2 < x\}$. Имеем:

$$\mathbb{P}\{\chi_n^2 < x\} = \mathbb{P}\{\xi_1^2 + \xi_2^2 + \dots + \xi_n^2 < x\}$$

Далее надо придумать такое множество $B \subset \mathbf{R}^n$, что

$$\mathbb{P}\{\xi_1^2 + \xi_2^2 + \dots + \xi_n^2 < x\} = \mathbb{P}(\xi_1, \xi_2, \dots, \xi_n) \in B\}$$

Очевидно, что

$$B = \{(x_1, x_2, \dots, x_n) : x_1^2 + x_2^2 + \dots + x_n^2 < x\}$$

Итак,

$$\begin{aligned}
&\mathbb{P}\{\chi_n^2 < x\} = \\
&= \int \dots \int_{x_1^2 + x_2^2 + \dots + x_n^2 < x} \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{x_1^2 + x_2^2 + \dots + x_n^2}{2}} dx_1 dx_2 \dots dx_n.
\end{aligned}$$

Далее мы перейдем к n -мерной сферической системе координат:

$$x_1 = r \cos \varphi_1,$$

$$x_2 = r \sin \varphi_1 \cos \varphi_2,$$

$$x_3 = r \sin \varphi_1 \sin \varphi_2 \cos \varphi_3,$$

...

$$x_{n-1} = r \sin \varphi_1 \sin \varphi_2 \dots \sin \varphi_{n-2} \cos \varphi_{n-1},$$

$$x_n = r \sin \varphi_1 \sin \varphi_2 \dots \sin \varphi_{n-2} \sin \varphi_{n-1},$$

У этой замены вычисляется якобиан, но мы попробуем эти вычисления обойти и использовать лишь тот факт, что переменная r входит в якобиан в $(n - 1)$ -й степени. Очевидно, что в новых координатах

$$x_1^2 + x_2^2 + \dots + x_n^2 = r^2.$$

Итак, мы приходим к следующему равенству:

$$\begin{aligned} & \int \dots \int_{x_1^2 + x_2^2 + \dots + x_n^2 < x} \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{x_1^2 + x_2^2 + \dots + x_n^2}{2}} dx_1 dx_2 \dots dx_n = \\ & = \int \dots \int_{r^2 < x} \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{r^2}{2}} C(\varphi_1, \varphi_2, \dots, \varphi_{n-1}) r^{n-1} dr d\varphi_1 \dots d\varphi_{n-1}. \end{aligned}$$

Далее мы интегралы по всем угловым переменным и все коэффициенты записываем как одну константу. В итоге мы приходим к равенству

$$\mathbb{P}\{\chi_n^2 < x\} = \text{Const} \int_{r^2 < x} e^{-\frac{r^2}{2}} r^{n-1} dr$$

Далее мы переходим к новой переменной $r^2 = u$ и еще раз при этом меняем константу:

$$\mathbb{P}\{\chi_n^2 < x\} = \text{Const} \int_{u < x} e^{-\frac{u}{2}} u^{\frac{n-2}{2}} du$$

Итак, $p_{\chi_n^2}(x) = \text{Const} e^{-x/2} x^{(n-2)/2}$. Проинтегрировав по x , мы определим константу:

$$\int_0^\infty \text{Const} e^{-x/2} x^{(n-2)/2} dx = 1, \text{Const} = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)}.$$

(Так как случайная величина χ^2 неотрицательна, $p_{\chi_n^2}(x) = 0$ при $x < 0$). Таким образом,

$$p_{\chi_n^2}(x) = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} e^{-x/2} x^{(n-2)/2} I_{(0, \infty)}(x).$$

Распределение Коши – это распределение случайной величины $\frac{\xi}{\eta}$, где ξ и η – независимые $N(0, 1)$ -нормальные случайные величины. Доказывается, что

$$p_{\xi/\eta}(x) = \frac{1}{\pi(1+x^2)}.$$

Доказательство проводится так же, как и выше, с использованием замены $y = u/v$, $z = v$. Нужно продифференцировать по x интеграл

$$\begin{aligned} & \int \int_{u/v < x} \frac{1}{2\pi} e^{-\frac{u^2+v^2}{2}} dudv = \int \int_{y < x}^{\infty} \frac{1}{\pi} e^{-\frac{z^2(1+y^2)}{2}} z dy dz = \\ & = \int_{y < x} \frac{1}{\pi(1+y^2)} dz. \end{aligned}$$

Замечание. Можно показать, что то же распределение имеет случайная величина $\frac{\xi}{|\eta|}$. Сведение этой задачи к предыдущей основано на симметричности распределений независимых случайных величин ξ и η . Легко показать, что если ξ имеет симметричное распределение, а ε – независимая от ξ случайная величина, принимающая значения $+1$ и -1 с вероятностью

$1/2$ (случайный знак, кстати, $\varepsilon = 1/\varepsilon$), то распределения случайных величин ξ , $\varepsilon \cdot \xi$, $\varepsilon \cdot |\xi|$ совпадают. Пусть ε не зависит от ξ и η . Тогда распределения $\frac{\xi}{|\eta|}$, $\frac{\varepsilon \xi}{|\eta|}$, $\frac{\xi}{\varepsilon |\eta|}$, $\frac{\xi}{\eta}$ совпадают.

Исключительно важным для статистики является *распределение Стьюдента* — распределение случайной величины

$$t_n = \frac{\xi}{\sqrt{\frac{1}{n}(\xi_1^2 + \dots + \xi_n^2)}},$$

где ξ, ξ_1, \dots, ξ_n независимые $N(0, 1)$ -нормальные случайные величины. Для вывода мы можем использовать функцию плотности распределения и записав совместную плотность независимых случайных величин ξ и χ_n^2 , найти, рассуждая как и при выводе распределения Коши, плотность случайной величины $\frac{\xi}{\sqrt{\frac{1}{n}\chi_n^2}}$. Мы этот вывод проводить здесь не будем.

9. Независимые случайные векторы. Введем независимые двумерные векторы. Для многомерных векторов (которые могут иметь даже разные размерности) определение независимости аналогично.

Определение. Случайные векторы $(\xi_1, \eta_1), (\xi_2, \eta_2), \dots, (\xi_n, \eta_n)$ называются независимыми, если для любых борелевских множеств $B_i \subset \mathbf{R}^2$ ($i \leq n$) события $\{(\xi_i, \eta_i) \in B_i\}$ независимы в совокупности.

Упражнения. а) Сформулируйте определение независимости случайных векторов (ξ_1, η_1) и (ξ_2, η_2, ζ_2) .

б) Покажите, что из независимости случайных векторов (ξ_1, η_1) и (ξ_2, η_2) следует независимость пар случайных величин ξ_1 и ξ_2 , ξ_1 и η_2 , η_1 и ξ_2 , η_1 и η_2 . Какие четыре пары двумерных борелевских множеств B_1 и B_2 нужно рассмотреть?

10. Как сконструировать набор (ξ_i) ($i \leq n$) независимых случайных величин с заданными распределениями \mathbb{P}_i ? Для этого надо использовать известную вам теорему Фубини. Мы представляем Ω как \mathbf{R}^n , на нем задаем как σ -алгебру событий σ -алгебру $\mathfrak{B}(\mathbf{R}^n)$ всех борелевских множеств, а в качестве вероятности \mathbb{P} на этой σ -алгебре вводим произведение вероятностей соотношением

$$\mathbb{P}\left(\prod_{i=1}^n B_i\right) = \prod_{i=1}^n \mathbb{P}_i(B_i).$$

Согласно теореме Фубини эта мера продолжается с произведений одномерных борелевских множеств на произвольные борелевские множества в \mathbf{R}^n . Случайные исходы — это точки n -мерного пространства, а случайные величины ξ_i задаются соотношениями $\xi_i(x_1, \dots, x_n) = x_i$.

11. Независимые приращения. В теории случайных процессов часто возникают такие наборы случайных величин $\xi_1, \xi_2, \dots, \xi_n$, что $\xi_1, \xi_2 - \xi_1, \dots, \xi_n - \xi_{n-1}$ независимы и имеют функции плотности p_1, p_2, \dots, p_n . Как выразить через них совместную плотность $p_{\xi_1, \xi_2, \dots, \xi_n}$?

Пусть борелевское множество $B \subset R^n$ произвольно, а \tilde{B} связано с ним соотношением:

$$\tilde{B} = \{(y_1, y_2, \dots, y_n) : (y_1, y_1 + y_2, \dots, y_1 + \dots + y_n) \in B\}.$$

Тогда

$$\mathbb{P}\{(\xi_1, \xi_2, \dots, \xi_n) \in B\} = \mathbb{P}\{(\xi_1, \xi_2 - \xi_1, \dots, \xi_n - \xi_{n-1}) \in \tilde{B}\}$$

$$= \int \dots \int_{\tilde{B}} p_1(y_1) \cdots p_n(y_n) dy_1 \cdots dy_n = \int \dots \int_B p_1(x_1) p_2(x_2 - x_1) \cdots p_n(x_n - x_{n-1}) dx_1 \cdots dx_n.$$

Отсюда и из определения совместной плотности следует

$$p_{\xi_1, \xi_2, \dots, \xi_n}(x_1, \dots, x_n) = p_1(x_1) p_2(x_2 - x_1) \cdots p_n(x_n - x_{n-1}).$$

Заметим, что якобиан использованного нами преобразования $(x_1, \dots, x_n) = (y_1, y_1 + y_2, \dots, y_1 + \dots + y_n)$ равен 1.

12. Симметризация. Используются два метода перехода от произвольной случайной величины ξ_n к случайной величине с симметричным распределением:

i) Вводится случайная величина η_n , которое имеет то же распределение, что и ξ_n и не зависит от нее. Тогда случайная величина $\xi_n^s = \xi_n - \eta_n$ имеет симметричное распределение.

ii) Вводится случайный знак ε_n , который имеет распределение $\mathbb{P}\{\pm 1\} = 1/2$ и не зависит от ξ_n . Тогда случайная величина $\varepsilon_n \cdot \xi_n$ имеет симметричное распределение.

Замечания. 1. Процедура симметризации обычно применяется к последовательностям случайных величин, этим объясняется обозначение. 2. Случайная величина $|\varepsilon_n \cdot \xi_n|$ имеет то же распределение, что и $|\xi_n|$. Другая симметризация, ξ_n^s , такого свойства не имеет.

Упражнение. Докажите все сформулированные в данном пункте утверждения.

Упражнения 13. а. Пусть n случайных величин ξ_i независимы и имеют одно и то же экспоненциальное распределение с функцией плотности $p(x) = e^{-x} I(x)$. Записать совместную функцию плотности p_{ξ_1, \dots, ξ_n} .

б. Пусть ξ имеет стандартное нормальное распределение, $\mathbb{P}\{\xi = \pm 1\} = \frac{1}{2}$, случайные величины ε и ξ независимы. Покажите, что случайные величины ξ , $\varepsilon \cdot \xi$ и $\varepsilon \cdot |\xi|$ имеют одинаковое распределение.

с. Пользуясь таблицами квантилей распределения Стьюдента найдите x , для которого $\mathbb{P}\{|\varphi| > x\} = 0.1$, где φ имеет распределение Коши.

д. Почему в таблице распределения χ_n^2 отсутствует $n = 1$?

Решение задач. 1.

Некоторые хотят решать задачи по теории вероятностей по формулам, использовать какую-нибудь формулу, подставить как-нибудь в эту формулу числа из условия задачи, и этим ограничиться. Я хочу вас предостеречь, ваша задача на практических занятиях значительно сложнее. В задачах на классическую модель вы должны **придумать** пространство элементарных исходов, **убедиться** (убедить себя), что элементы этого пространства равновероятны. Часто в этом помогает воображение, вы должны подумать, что же происходит в предложенном в задаче 'эксперименте', и представить себе все возможные исходы. При этом мы вынуждены идеализировать ситуацию, например, все карты равноправны, и их не раздает карточный шулер, орел и решка равноправны (хотя картинки разные, и стороны монеты незначительно отличаются), отталкиваясь от этих простейших симметрий, мы строим по ним более сложные симметрии. Например, если результаты эксперимента записываются в виде (a, b) , где возможны и явно равноправны все варианты $a \in A$ и $b \in B$, то мы можем считать равными вероятности всех исходов $(a_1, b_1), \dots, (a_N, b_N)$, где $N = |A| \times |B|$. Если результаты экспериментов более сложные — (a_1, \dots, a_n) где $a_i \in A_i$, то, как я вам уже говорил, число элементарных исходов считается по формуле $|A_1| \times |A_2| \times \dots \times |A_n|$, а если все $A_i = A$, то $|A|^n$. Например, в некоторых задачах вы подбрасываете монету несколько раз — n , но результат подбрасывания вы считаете не n экспериментами, а одним сложным экспериментом. В этом случае возникает соблазн уменьшить число элементарных исходов и считать одинаковыми наборы, состоящие из одинаковых элементов (вне зависимости от порядка этих элементов). Как правило, это некорректно. Вы заменяете множество на классы эквивалентности, не учитывая то обстоятельство, что классы эквивалентности состоят из разного числа элементов, то есть заведомо не равновероятны. Лишь только в квантовой физике, в статистике Бозе – Эйнштейна это принято, и думаю, что эту статистику физики много раз проверяли на эксперименте, прежде чем ей поверить. Иногда ситуация проведения таких экспериментов усложняется, результат одного частного эксперимента влияет на исход следующего эксперимента, например, они не могут повторяться. Это происходит, когда вы используете модели, основанные на выборках без возвращения. У вас появляется вполне корректная возможность рассматривать вместо упорядоченных выборок неупорядоченные. В этом случае нужно следить, чтобы и при подсчете числа благоприятству-

ющих исходов вы использовали выборки того же вида. В таких задачах возможны нюансы. Например, в одной задаче можно рассматривать число разбиений четырехэлементного множества на две пары. Каждое разбиение задает выбор пары, число способов такого выбора - 6. Но каждое разбиение можно задать выбором двух пар, поэтому разбиений вдвое меньше - 3. Имеются задачи, казалось бы, на неклассическую модель, так как в них бесконечное число исходов (например, мы бросаем монету не фиксированное заранее число раз). Однако часто в таких задачах можно использовать классическую модель, например, для вычисления вероятности каждого исхода надо строить свою конечную классическую модель и подсчитать таким способом вероятность данного исхода.

Похожие трудности возникают также при использовании геометрической модели. И здесь рекомендуется чертить для себя множество всех исходов.

Я как-то побывал на конференции в Тимишоаре (город, в котором служил армейским офицером Больяни, математик, независимо от Лобачевского создавший неевклидову геометрию). На улице Больяни имеется доска, на которой на нескольких языках, в том числе и на русском, написана фраза из письма Больяни: 'Из ничего я создал мир' (ничего - это отрицание пятого постулата).

При решении каждой вероятностной задачи вы должны фантазировать и каждый раз со-здавать новый мир - мир элементарных исходов, и попытаться разобраться, что же в этом мире может произойти.

Пример 1. Ведущий говорит, что в одной из трех шкатулок приз, и предлагает выбрать шкатулку. Играющий делает выбор, после чего ведущий, который знает, в какой из трех шкатулок находится приз, открывает одну из двух оставшихся шкатулок, показывает, что в ней ничего нет, и провокационно спрашивает: 'А может быть, Вы передумаете?'. Есть ли смысл для играющего передумывать?

Решение. Эта задача рассматривалась в средствах массовой информации. Первая мысль — какой смысл передумывать. Остались две неоткрытые шкатулки, которые по условиям задачи равноправны. Поэтому замена одной из них на другую не изменит вероятность. Оказывается, это неверно. Чтобы понять ситуацию, полезно вспомнить, что вероятность события — это предел частоты выпадения события. Ясно, что доля экспериментов, в которых выбранная играющим шкатулка содержит приз, равна примерно трети, и действия ведущего не могут в экспериментах невидимым способом убрать или добавить приз в выбранную шкатулку. Таким образом, вне зависимости от действий ведущего приз в выбранной играющим шкатулке находится лишь в третьей части экспериментов. В двух третях экспериментов приз находится в двух невыбраных шкатулках. В показанной ведущим шкатулке приза нет, значит, в двух третях экспериментов он находится в невыбранной играющим и не открытой ведущим шкатулке. Итак, если не передумывать, вероятность получить приз равна $\frac{1}{3}$, если передумать, эта вероятность равна $\frac{2}{3}$.

Разумеется, все это верно, лишь если ведущий обязан по правилам игры открыть шкатулку без приза. Если он этого делать не обязан, то он может вредничать и чаще показывать шкатулку в случае правильного первоначального выбора. Задача превращается в задачу из теории игр с вероятностными стратегиями двух игроков.

Пример 2. Четыре человека сдали шляпы. Гардеробщик был не в форме и отдал их наугад. Какова вероятность того, что ровно двое человек получили свои шляпы?

Решение. В этой задаче легко определяется пространство Ω . Оно состоит из множества всех перестановок четырех шляп (или их хозяев). Итак, $|\Omega| = 4!$. Теперь зададим множество благоприятствующих перестановок. Каждая из них однозначно задана, если заданы два счастливых получателя своих шляп. Чтобы получилась перестановка с нужным свойством, другие двое должны шляпами обменяться. Чтобы задать такую перестановку, достаточно задать пару. Число вариантов выбора пары равно C_4^2 . Ответ — $\frac{6}{24}$.

Пример 3. В карточной игре 'преферанс' у играющего в выбранной им козырной масти *Туз, Король, 7, 8*. Найти вероятность того, что у двух других игроков (у них по 10 карт) карты *Дама, Валет, 10, 9* распределились пополам.

Решение. При решении этой задачи много вариантов выбора вероятностной модели. Самым простым способом нам представляется нумерация 20 карт оставшихся игроков, причем такая, что первый игрок имеет карты от 1-й до 10-й, а второй игрок — от 11-й до 20-й. Элементарный исход — номера *Дамы*, *Вальта*, 10, 9. $|\Omega| = 20 \cdot 19 \cdot 18 \cdot 17$. Теперь найдем $|A|$. $A = A_{12} + A_{13} + A_{14} + A_{23} + A_{24} + A_{34}$, где A_{ij} — исходы, в которых первый игрок получил *Даму*, *Вальта*, а второй — 10, 9, где A_{ij} — исходы, в которых первый игрок получил *Даму*, 10, а второй — *Вальта*, 9, и т. д. Очевидно, что

$$|A_{12}| + |A_{13}| + |A_{14}| + |A_{23}| + |A_{24}| + |A_{34}|.$$

Поэтому $|A|$ равно числу $|A_{12}|$, умноженному на 6 — число пар двух карт из четырех, доставшихся первому игроку. Заметим, что произвольный $\omega \in A_{12}$ можно представить как

$$\omega = (\omega_1, \omega_2),$$

где ω_1 — номера *Дамы* и *Вальта*, доставшихся первому игроку, а ω_2 — номера 10 и 9, доставшихся второму первому игроку. Рассуждая так же, как и при подсчете $|\Omega|$, имеем: $|A| = 10 \cdot 9 \cdot 10 \cdot 9$.

Ответ: $6 \frac{10 \cdot 9 \cdot 10 \cdot 9}{20 \cdot 19 \cdot 18 \cdot 17}$.

Пример 4. Монета подбрасывается до тех пор, пока не выпадет герб. Найти вероятность того, что монета будет подбрасываться четное число раз.

Решение. Очевидно, что число элементарных исходов в этой задаче бесконечно, это — множество всех натуральных чисел. Но при вычислении числа $\mathbb{P}\{n\}$ мы используем временную модель, в которой монета бросается лишь n раз, и общее число элементарных исходов равно 2^n , но лишь один из них удовлетворяет условию {решка, решка, ..., решка, герб} (решка выпадала первые $n - 1$ раз). Итак, $\mathbb{P}\{n\} = 2^{-n}$, а искомая вероятность — сумма геометрической прогрессии

$$\frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \dots = \frac{1}{3}.$$

Пример 5. Два автобуса ходят с интервалом движения 10 мин. Какова вероятность уехать на одном из них за 5 минут?

Решение. В этой задаче мы используем геометрическую модель, элементарными исходами в которой являются пары чисел (x, y) , где x — время ожидания первого автобуса, y — время ожидания второго автобуса. $\mathbb{P}(A) = \frac{m(A)}{m(\Omega)}$, m — площадь. Нельзя признать эту модель очень разумной, так как время ожидания автобуса может и превысить 10 минут, мы ее используем лишь ввиду отсутствия всякой дополнительной информацией, за неимением лучшей модели. Такая модель лучше подходит для метро, где интервал движения выдерживается достаточно точно. Итак,

$$\Omega = \{(x, y) : 0 \leq x \leq 10, 0 \leq y \leq 10\}, A = \{(x, y) : 0 \leq x \leq 5 \text{ или } 0 \leq y \leq 5\}.$$

$$\mathbb{P}(A) = \frac{75}{100}.$$

Пример 6. Симметричная монета бросается до тех пор, пока не выпадет герб. Найти вероятность того, что она бросалась нечетное число раз.

Решение. Задача эквивалентна примеру 4, но здесь решение будет использовать условные вероятности. В этой задаче

$$\Omega = \{1, 2, 3, \dots\}, A = \{1, 3, 5, \dots\}.$$

Однако число исходов бесконечно, поэтому неприменима классическая модель (в самом деле, почему вероятность первого герба в первом испытании равна вероятности первого герба в десятом испытании). Тем не менее, классическая модель используется для подсчета $\mathbb{P}(\{n\})$, причем для каждого n — своя. Итак, чему равна вероятность $\mathbb{P}(\{n\})$. Вводим модель Ω_n , состоящую

из всех наборов гербов и решек длины n . $\Omega_n = 2^n$. Из этих наборов лишь один соответствует исходу $\{n\}$, — (решка, решка,...,решка, герб), в котором $n - 1$ решка, а на последнем месте герб. Итак,

$$\mathbb{P}(A) = \frac{1}{2^n}, \mathbb{P}(A) = \sum_{k=0}^{\infty} 2^{-2k+1} = \frac{2}{3}.$$

Любопытно, что эту задачу можно решить и без построения и анализа Ω , используя прием из задач о разорении игрока и о распаде атома:

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A|\text{герб в 1 бросании})\mathbb{P}(\text{герб в 1 бросании}) + \\ &+ \mathbb{P}(A|\text{решка в 1 бросании})\mathbb{P}(\text{решка в 1 бросании}). \end{aligned}$$

Заметим, что

$$\mathbb{P}(A|\text{решка в 1 бросании}) = \mathbb{P}(\bar{A}),$$

так как если в первом бросании выпадет решка, то второе бросание становится первым после этой решки, каждое четное бросание становится нечетным, \bar{A} превращается в A , $\mathbb{P}(A|\text{герб в 1 бросании}) = 1$. Итак,

$$\mathbb{P}(A) = 1 \cdot \frac{1}{2} + (1 - \mathbb{P}(A)) \frac{1}{2}, \mathbb{P}(A) = \frac{2}{3}.$$

Пример 7. В столе 7 ящиков, и в нем находится письмо с вероятностью p , при чем вероятности нахождения письма в каждом из ящиков равны. Открыли 6 ящиков, письма нет. Какова вероятность того, что письмо находится в седьмом ящике. Эта замечательная задача заимствована из задачника Л.Д. Мешалкина (МГУ). Сама задача наталкивает на неправильный ответ — p . Мой опыт показывает, что даже после демонстрации решения бывает трудно убедить в неверности этого ответа нематематиков (механиков, физиков, причем даже очень талантливых, даже если они работают на математических кафедрах и преподают математику). В то же время студенты мехмата быстро понимают, в чем дело. Для объяснения неверности напрашивавшегося ответа лучше всего рассмотреть частный случай: $p = \frac{7}{12}$, дробное p легко моделируется. Давайте добавим к столу еще один стол с точно такими же 5 ящиками и будем считать, что письмо находится в одном из 12 ящиков с вероятностью 1. Письма нет в шести ящиках, значит, оно находится в одном из оставшихся 6 ящиков. Седьмой ящик нашего стола — это один из оставшихся шести ящиков, поэтому, в нашем случае ответ должен быть равен $\frac{1}{6}$

Решение. Представим интересующее нас число как

$$\begin{aligned} \mathbb{P}(\{\text{письмо в седьмом ящике}\} | \{\text{письма нет в первых шести ящиках}\}) &= \\ \frac{\mathbb{P}(\{\text{письмо в седьмом ящике}\} \{\text{письма нет в первых шести ящиках}\})}{\mathbb{P}(\{\text{письма нет в первых шести ящиках}\})}. \\ \mathbb{P}(\{\text{письма нет в первых шести ящиках}\}) &= 1 - \frac{6}{7}p, \\ \mathbb{P}(\{\text{письмо в седьмом ящике}\} \{\text{письма нет в первых шести ящиках}\}) &= \frac{1}{7}p. \end{aligned}$$

Пример 8. Имеются две урны, в одной 5 белых и 2 черных шара, в другой — 3 белых и 4 черных шара. Наугад (с вероятностью $\frac{1}{2}$) выбрали урну и вытащили из нее шар, который оказался белым. Этот шар положили на место и снова из той же урны вытащили шар. Какова вероятность, что он окажется белым?

Решение. Казалось бы, правильный ответ:

$$\frac{5}{7} \cdot \frac{1}{2} + \frac{3}{7} \cdot \frac{1}{2} = \frac{4}{7}.$$

Очевидно, что мы уже выбрали урну с вероятностью $\frac{1}{2}$, и ситуация не изменилась от того, что мы из нее вытащили шар. Но это неверно, и чтобы убедиться в неверности такого решения,

полезно рассмотреть крайний случай — во второй урне 7 черных шаров. В этой ситуации ясно, что мы выбрали первую урну, вероятность выбрать из нее белый шар равна не $\frac{4}{7}$, а

$$\frac{5}{7} = \frac{5}{7} \cdot 1 + \frac{3}{7} \cdot 0.$$

Итак, результат первого эксперимента меняет вероятности выбора урн, они становятся условными, и вычисляются по формуле Байеса.

$$\mathbb{P}(1 \text{ урна} | \text{белый шар}) = \frac{\frac{5}{7} \cdot \frac{1}{2}}{\frac{5}{7} \cdot \frac{1}{2} + \frac{3}{7} \cdot \frac{1}{2}} = \frac{5}{8}.$$

Далее в формуле полной вероятности мы используем вместо первоначальной $\frac{1}{2}$ новые вероятности выбора урн (которые так и хочется назвать виртуальными):

$$\frac{5}{7} \cdot \frac{5}{8} + \frac{3}{7} \cdot \frac{3}{8} = \frac{34}{56}.$$

Процесс выбора шаров может продолжаться, и после каждого последующего результата мы должны пересчитать вероятность выбора урны, используя при этом вероятность, сосчитанную после предыдущего результата.

Пример 9. Два игрока играют в разную силу, у первого вероятность выигрыша партии равна $\frac{2}{5}$, у второго — $\frac{3}{5}$. Результат игры определяется с учетом гандикапа — первому для победы надо выиграть три партии, второму — две. Результат будет известен через четыре партии. У кого из игроков вероятность выигрыша больше?

Решение. Событие 'выигрыш второго игрока' представляется в виде суммы трех несовместных событий $A_2 + A_3 + A_4$ (A_i — событие, означающее выигрыш i партий вторым игроком). Вероятность каждого из них считается по формуле Бернулли: $\mathbb{P}(A_i) = C_4^i \left(\frac{2}{5}\right)^i \left(\frac{3}{5}\right)^{4-i}$.

Примечание. Может появиться соблазн после выигрыша второй партии вторым игроком больше не играть. Но тогда общее число партий изменится и разные события A_i будут относиться к разным моделям. Такой путь решения лишь усложнит решение.

Пример 10. Посмотрите задачу о распаде атома.

Пример 11. Посмотрите задачу об электрической цепи.

Пример 12. Вероятности отметок (с учетом неявки и отказа) по философии и теории вероятностей являются следующими (ξ — отметка по философии, η — отметка по теории вероятностей).

$$\begin{aligned} \mathbb{P}\{\xi = 2, \eta = 2\} &= \frac{1}{10}, \mathbb{P}\{\xi = 2, \eta = 3\} = \frac{1}{20}, \mathbb{P}\{\xi = 2, \eta = 4\} = \frac{1}{20}, \mathbb{P}\{\xi = 2, \eta = 5\} = \frac{1}{20}, \\ \mathbb{P}\{\xi = 3, \eta = 2\} &= \frac{1}{10}, \mathbb{P}\{\xi = 3, \eta = 3\} = \frac{1}{20}, \mathbb{P}\{\xi = 3, \eta = 4\} = \frac{1}{20}, \mathbb{P}\{\xi = 3, \eta = 5\} = \frac{1}{20}, \\ \mathbb{P}\{\xi = 4, \eta = 2\} &= \frac{1}{20}, \mathbb{P}\{\xi = 4, \eta = 3\} = \frac{1}{10}, \mathbb{P}\{\xi = 4, \eta = 4\} = \frac{1}{10}, \mathbb{P}\{\xi = 4, \eta = 5\} = \frac{1}{20}, \\ \mathbb{P}\{\xi = 5, \eta = 2\} &= \frac{1}{20}, \mathbb{P}\{\xi = 5, \eta = 3\} = \frac{1}{20}, \mathbb{P}\{\xi = 5, \eta = 4\} = \frac{1}{20}, \mathbb{P}\{\xi = 5, \eta = 5\} = \frac{1}{20}. \end{aligned}$$

Являются ли эти оценки независимыми случайными величинами.

Решение. Нет, так как например,

$$\mathbb{P}\{\xi = 2\} = \frac{1}{10} + \frac{1}{20} + \frac{1}{20} + \frac{1}{20} = \frac{5}{20},$$

$$\mathbb{P}\{\eta = 2\} = \frac{1}{10} + \frac{1}{10} + \frac{1}{20} + \frac{1}{20} = \frac{6}{20},$$

$$\mathbb{P}\{\xi = 2, \eta = 2\} = \frac{1}{10} \neq \frac{5}{20} \cdot \frac{6}{20}.$$

Доказывать независимость этих величин (если она есть) значительно сложнее, нужно проверить независимость 12 пар событий (почему не 16?).

Пример 13. В предыдущем примере найдите распределение случайной величины ξ .

Пример 14. Совместная плотность $p_{\xi, \eta}$ случайных величин ξ и η равна 1 в параллелограмме $ABCD$ с вершинами $A = (0, 0)$, $B = (1, 0)$, $C = (1, 1)$, $D = (1, 2)$, и равна 0 вне параллелограмма. а) Найти плотность p_ξ . б) Известно, что случайные величины $\xi + c\eta$, η независимы. Найти константу c .

Решение. а) Легко подсчитать функцию распределения, S — площадь:

$$F_\xi(t) = S(\{(x, y) : x < t\} \cap ABCD) = \begin{cases} t^2/2, & \text{если } 0 \leq t \leq 1; \\ 1 - (2-t)^2/2, & \text{если } 1 \leq t \leq 2. \end{cases}$$

Чтобы найти плотность, надо продифференцировать. Задачу б) решите сами.

Пример 15. Распределение случайного вектора (ξ, η) таково, что вся вероятность находится на ломаной L , соединяющей точки $(0, 0), (1, 1), (2, 0)$, причем вероятность попасть в часть ломаной пропорциональна длине l этой части. Найти распределение случайной величины ξ .

Решение. $F_\xi(t) = \frac{l(\{(x, y) : x < t\} \cap L)}{l(L)} = t/2$. Плотность равна $\frac{1}{2} \cdot I_{[0, 2]}$.

Пример 16. Случайные величины ξ , η независимы и имеют показательное распределение с функцией плотности $p(x) = e^{-x} I_{[0, \infty)}(x)$. Доказать, что случайные величины $\xi + \eta$, ξ/η независимы.

Решение. Чтобы доказать независимость в данной ситуации существования совместной функции плотности достаточно проверить равенство

$$p_{\xi+\eta, \xi/\eta}(x, y) = p_{\xi+\eta}(x)p_{\xi/\eta}(y).$$

Для этого нужно вспомнить определение совместной функции плотности через вероятности событий: для любого boreлевского множества $B \subset \mathbf{R}^2$ должно быть верно

$$\mathbb{P}\{(\xi + \eta, \xi/\eta) \in B\} := \int \int_B p_{\xi+\eta, \xi/\eta}(x, y) dx dy. \quad (*)$$

Мы не знаем совместную функцию плотности $p_{\xi+\eta, \xi/\eta}$, но зато можем записать, используя условия задачи, совместную функцию плотности:

$$p_{\xi, \eta}(u, v) = e^{-u} I_{[0, \infty)}(u) e^{-v} I_{[0, \infty)}(v).$$

Разумеется, нужно вспомнить еще раз определение совместной функции плотности: для любого boreлевского множества $\tilde{B} \subset \mathbf{R}^2$

$$\mathbb{P}\{(\xi, \eta) \in \tilde{B}\} := \int \int_{\tilde{B}} p_{\xi, \eta}(u, v) du dv. \quad (**)$$

Связь между $(*)$ и $(**)$ можно осуществить лишь через вероятности событий. За счет выбора \tilde{B} нужно добиться того, чтобы

$$\{(\xi + \eta, \xi/\eta) \in B\} = \{(\xi, \eta) \in \tilde{B}\}.$$

Вспомним, что каждое множество определяется правилом, определяющим точки множества. Очевидно, что

$$\tilde{B} = \{(u, v) : (u + v, u/v) \in B\}.$$

Итак,

$$\mathbb{P}\{(\xi + \eta, \xi/\eta) \in B\} = \int \int_{\{(u, v) : (u + v, u/v) \in B\}} e^{-(u+v)} I_{[0, \infty)}(u) I_{[0, \infty)}(v) du dv.$$

После напрашивающейся замены переменных $x = u + v$, $y = u/v$, выразив u , v через x , y и вычислив якобиан, получим

$$\int \int_B p_{\xi+\eta, \xi/\eta}(x, y) dx dy = \int \int_B e^{-x} I_{[0, \infty)}(x) I_{[0, \infty)}(y) \frac{x}{(y+1)^2} dx dy,$$

что влечет независимость $\xi + \eta$, ξ/η . Наши выкладки содержат и логический элемент — мы проверили, что переменные x , y неотрицательны тогда и только тогда, когда неотрицательны u , v , и исходя из этого, преобразовали индикаторы.

Вопросы, на которые нужно отвечать сразу:

1) Определение независимых событий. Нужно понимать, что верно и обратное, если вероятность произведения равна произведению вероятностей, то эти события независимы, даже если на первый взгляд кажутся зависимыми. Верно и противоположное, если нужного равенства вероятностей нет, то события зависимы.

2) События, независимые в совокупности. Вероятность произведения равна произведению вероятностей для любого поднабора.

3) Несовместные события, их нельзя путать с независимыми. Нужно понимать, что для несовместных A и B имеет место $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ (мы пишем, используя дополнительную нагрузку на символ, $\mathbb{P}(A + B) = \mathbb{P}(A) + \mathbb{P}(B)$). Это частный случай формулы сложения, используется при решении задач. Для независимых событий верна другая формула (ее помнить не обязательно) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)$.

Итак, формула $\mathbb{P}(A + B) = \mathbb{P}(A) + \mathbb{P}(B)$ будет использоваться в самых разных вариантах. Например, $\mathbb{P}(A) = \mathbb{P}(\Omega) - \mathbb{P}(\bar{A})$ — частный случай формулы $\mathbb{P}(A) = \mathbb{P}(A + B) - \mathbb{P}(B)$. Итак, для вычисления (или оценки сверху) вероятности события его нужно разбить на несколько несовместных событий, каждое из которых легче вычислить (или оценить).

4) Определение условной вероятности.

5) Вычисление вероятности попадания случайного вектора в множество в дискретном и непрерывном случаях. Приведем алгоритм для двумерного случайного вектора (ξ, η) . В дискретном случае знание распределения означает знание всех вероятностей

$$\mathbb{P}\{(\xi, \eta) = (x_i, y_i)\} = p_i, \text{ где } \sum_i p_i = 1.$$

Итак, для вычисления $\mathbb{P}\{(\xi, \eta)\}$, надо сложить все числа p_i , для которых $(x_i, y_i) \in B$. Например, для вычисления $\mathbb{P}\{\xi = x\}$ надо сложить все числа p_i , для которых $x_i = x$ (в этом случае $B = \{x\} \times \mathbf{R}$).

Аналогично, в непрерывном случае мы должны взять интеграл по B от совместной функции плотности. Обратно, это равенство является определением совместной функции плотности. Поэтому в двумерном случае функция плотности ξ задается интегралом (точнее, может быть задана интегралом, так как функция плотности задана с точностью до значений на множестве лебеговой меры нуль)

$$p_\xi(x) = \int_{\mathbf{R}} p_{\xi, \eta}(x, y) dy.$$

Действительно, при таком задании мы будем иметь тождество для любого борелевского B :

$$\begin{aligned} \mathbb{P}\{\xi \in B\} &= \int_B p_\xi(x) dx = \\ &= \mathbb{P}\{(\xi, \eta) \in B \times \mathbf{R}\} = \int_{B \times \mathbf{R}} p_{\xi, \eta}(x, y) dx dy. \end{aligned}$$

6) Определение независимости случайных величин: ξ и η :

$$\mathbb{P}\{(\xi, \eta) \in B_1 \times B_2\} = \mathbb{P}\{\xi \in B_1\} \cdot \mathbb{P}\{\eta \in B_2\}.$$

В дискретном случае нужно проверить, что для всех возможных значений x и y случайных величин ξ и η имеет место равенство

$$\mathbb{P}\{\xi = x, \eta = y\} = \mathbb{P}\{\xi = x\} \cdot \mathbb{P}\{\eta = y\}. \quad (1)$$

Определение получается сложением этих равенств по всем возможным значениям $x \in B_1$ и $y \in B_2$ случайных величин ξ и η . В непрерывном случае для независимости необходимо и достаточно разложение

$$p_{\xi, \eta}(x, y) = p_\xi(x) \cdot p_\eta(y) \quad (2)$$

для всех x и y , за исключением множества лебеговой меры нуль.

7) Как записывается совместная функция плотности независимых случайных величин $\xi_1, \xi_2, \dots, \xi_n$, имеющих одну и ту же функцию плотности p .

8) Вычисление среднего в дискретном случае

$$\mathbb{E}\xi = \sum_i x_i \mathbb{P}\{\xi = x_i\},$$

суммирование по всем возможным значениям x_i случайной величины ξ . Соответственно, если заданы вероятности значений случайного вектора (ξ, η) , то

$$\mathbb{E}\xi = \sum_i x_i \mathbb{P}\{\xi = x_i, \eta = y_i\},$$

суммирование по всем возможным значениям (x_i, y_i) случайного вектора (ξ, η) .

9) Вычисление среднего в непрерывном случае

$$\mathbb{E}\xi = \int_{\mathbf{R}} xp_\xi(x) dx.$$

Соответственно, если задана совместная плотность случайного вектора (ξ, η) , то

$$\mathbb{E}\xi = \int_{\mathbf{R}} \int_{\mathbf{R}} xp_{\xi, \eta}(x, y) dx dy.$$

10) Аддитивность среднего, константа выносится из под знака среднего, среднее константы равно этой константе, в частности $\mathbb{E}(\mathbb{E}\xi) = \mathbb{E}\xi$ (среднее значение — это константа).

11) Среднее произведения независимых случайных величин равно произведению средних.

12) Два определения дисперсии:

$$\mathbb{D}\xi = \mathbb{E}(\xi - \mathbb{E}\xi)^2 = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2.$$

13. Свойства дисперсии — дисперсия не меняется при сдвиге случайной величины на константу ($\mathbb{D}\xi = \mathbb{D}(\xi + c)$), при умножении случайной величины на константу ее дисперсия умножается на квадрат константы, и, наконец, дисперсия суммы независимых случайных величин равна сумме дисперсий.

14) Параметры нормального распределения: m — среднее, σ^2 — дисперсия.

15) Определение функции распределения, ее связь с функцией плотности.

Требование отвечать на эти вопросы сразу вынужденно. Все это многократно используется, и если не воспринимать эти вещи как очевидные, то вы не будете успевать понимать лекции в следующем семестре. Так что рекомендуется несколько раз все это повторить, причем в конце ваших занятий, когда вы устали. Спрашиваться это будет в конце экзамена.

§7. Характеристики случайных величин

1. Среднее значение или математическое ожидание.

Еще раз о классической модели и общей ситуации.

Наша дальнейшая задача — научиться интегрировать случайные величины.

Наличие классической модели означает, что мы знаем элементарные исходы, т.е. мы полностью владеем ситуацией. Мы не знаем, какой элементарный исход сейчас произойдет. Во многих случаях такой информацией у нас нет. Например, изучая длину растений (ее распределение согласно многочисленным экспериментам нормально), мы не знаем и не можем учесть всех факторов, от которых зависит длина растения. Грубо говоря, элементарный исход должен был бы представлять очень длинную цепочку типа (99 капель дождя, сухо с температурой 12%, кусочек навоза в 1 грамм, 73 капли дождя,...). Итак, пространство элементарных исходов нам, как правило, неизвестно.

Нам известны два вида интеграла — Римана и Лебега. Процедура вычисления интеграла Римана явно использует структурные свойства числовой прямой — например, порядок, арифметические операции и т.д. Благодаря этому, вводится операция дифференцирования, а для вычисления интеграла используется обратная операция. Мы только что поняли, что в пространстве всех элементарных исходов может не оказаться никакой понятной нам структуры. Поэтому, естественно попытаться использовать здесь интеграл Лебега, для введения которого нужно лишь одно понятие — мера.

\mathbb{E} определяется для дискретной случайной величины $\sum_n a_n I_{\xi=a_n}$ равенством

$$\mathbb{E} \sum_n a_n I_{\{\xi=a_n\}} =: \sum_n a_n \mathbb{P}\{\xi = a_n\}.$$

Это определение корректно, если ряд сходится абсолютно. Если ряд расходится, то среднее значение не существует.

Это определение имеет ряд свойств интеграла, в частности, $|\xi| \leq C$ влечет $|\mathbb{E}(\xi)| \leq \mathbb{E}C = C$. Отсюда немедленно следует, что если последовательность случайных величин ξ_n сходится равномерно к ξ (напомню, это означает, что $\sup |\xi - \xi_n| \rightarrow 0$, а это влечет $\sup |\xi_m - \xi_n| \rightarrow 0$, если $m, n \rightarrow \infty$). Но тогда $|\mathbb{E}(\xi_m - \xi_n)| \leq \sup |\xi_m - \xi_n| \rightarrow 0$, то есть последовательность $\mathbb{E}\xi_n$ является фундаментальной, а следовательно, имеет предел, который мы обозначаем через $\mathbb{E}\xi$. Итак, понятие среднего мы можем распространить с дискретных случайных величин на пределы равномерно сходящихся последовательностей таких величин. Нам остается лишь представить любую случайную величину в виде равномерного предела дискретных случайных величин. Это делается просто. С каждой случайной величиной ξ мы связываем случайные величины

$$\xi^{(n)} = \sum_{k \in \mathbf{Z}} \frac{k}{n} I_{\left\{ \xi \in \left[\frac{k}{n}, \frac{k+1}{n} \right) \right\}}.$$

Очевидно, что $\xi^{(n)}$ равномерно сходится к ξ , более того, $|\xi^{(n)} - \xi| < 1/n$.

Замечание. Равномерную сходимость ξ_n к ξ можно было бы назвать сходимостью в известном вам пространстве $L_\infty(\Omega, \mathcal{A}, \mathbb{P})$. Но сами случайные величины ξ_n и ξ не обязательно принадлежат пространству $L_\infty(\Omega, \mathcal{A}, \mathbb{P})$ (они принадлежат $L_1(\Omega, \mathcal{A}, \mathbb{P})$). Однако их разности принадлежат $L_\infty(\Omega, \mathcal{A}, \mathbb{P})$ и у них можно вычислять норму в этом пространстве.

2. Свойства среднего

те же, что и у интеграла.

1. $\mathbb{E}(Const) = Const$.
2. $\mathbb{E}(a\xi + b\eta) = a\mathbb{E}\xi + b\mathbb{E}\eta$

для всех констант a, b и всех случайных величин ξ, η , имеющих среднее. В частности, среднее суммы равно сумме средних.

3. $\xi \leq \eta$ влечет $\mathbb{E}\xi \leq \mathbb{E}\eta$, в частности,

4. Если $|\xi| \leq c$, где c — константа, то $|\mathbb{E}\xi| \leq c$. Если ξ неотрицательна, $\mathbb{E}\xi = 0$, то $\mathbb{P}\{\xi = 0\} = 1$.

Замечание. Легко видеть, что введенный нами интеграл Лебега – это и интеграл $\int_{\Omega} \xi d\mathbb{P}$, который в теории вероятностей чаще всего обозначается $\mathbb{E}\xi$, но в ряде учебников обозначается также $\mathbb{M}\xi$. Используемые термины — *среднее значение* случайной величины ξ или *математическое ожидание* ξ .

Очень полезное замечание. Из свойства 4 следует, что равномерная сходимость случайных величин влечет сходимость средних (то есть, интегралов). Это позволяет во многих случаях сводить доказательства разного рода утверждений для среднего значения случайных величин к случаю, когда эта случайная величина является дискретной. В этом случае среднее значение оказывается конечной линейной комбинацией (или бесконечной линейной комбинацией, то есть суммой ряда) индикаторов событий. Поэтому для доказательства теоремы для случайных величин общего вида часто бывает достаточно доказать эту теорему для индикаторов, а потом перейти к линейным комбинациям индикаторов и их пределам. Впрочем, ниже мы имеем пример ситуации, когда равномерная сходимость такому переходу не помогает.

3. Формула замены переменной ($\omega \in \Omega$ заменяется на $x \in \mathbf{R}$).

Для дискретного типа распределения формула для среднего уже имеется. Теперь я хочу доказать формулу для среднего в случае непрерывного типа распределения:

$$\mathbb{E}\xi = \int_{\mathbf{R}} xp_{\xi}(x)dx,$$

которую мы уже использовали. Но мне будет удобнее доказать более общую формулу:

$$\mathbb{E}g(\xi_1, \dots, \xi_n) = \int_{\mathbf{R}^n} g(x_1, \dots, x_n)p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)dx_1 \dots dx_n, \quad (1)$$

где функцию g можем считать измеримой, хотя для приложений достаточно считать кусочно непрерывной, но при этом мы предполагаем, что наш n -мерный интеграл сходится абсолютно (в противном случае среднее не существует).

Дело в том, что функцию g мы также можем представить как равномерный предел дискретных функций

$$g^{(n)} = \sum_{k \in \mathbf{Z}} \frac{k}{n} I_{\left\{ g \in \left[\frac{k}{n}, \frac{k+1}{n} \right) \right\}}.$$

Но тогда $g^{(n)}(\xi_1, \dots, \xi_n)$ сходится равномерно к $g^{(n)}(\xi_1, \dots, \xi_n)$. Поэтому достаточно проверить формулу (1) для $g^{(n)}(\xi_1, \dots, \xi_n)$, а потом перейти к пределу. Имеем по определению среднего для дискретных случайных величин:

$$\begin{aligned} \mathbb{E}g^{(n)}(\xi_1, \dots, \xi_n) &= \sum_{k \in \mathbf{Z}} \frac{k}{n} \mathbb{P}_{\xi_1, \dots, \xi_n} \left\{ g^{(n)}(x_1, \dots, x_n) = \frac{k}{n} \right\} = \\ &= \sum_{k \in \mathbf{Z}} \frac{k}{n} \int_{\substack{p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \frac{k}{n}}} p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \int_{\mathbf{R}^n} g^{(n)}(x_1, \dots, x_n) p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) dx_1 \dots dx_n. \end{aligned}$$

В чем смысл доказанного утверждения. В анализе он называется *совпадением интегралов Римана и Лебега для функций, интегрируемых по Риману*. Доказательство чрезвычайно прозрачно. Мы приближаем функцию ступенчатой, пользуемся совпадением интегралов Римана и Лебега для ступенчатых функций, а потом переходим к пределу, устремляя ступенчатую функцию к интегрируемой по Риману. Польза от утверждения для теории вероятностей бесспорна

— мы умеем вычислять интеграл Римана, пользуясь техникой первообразных, и применяем эту технику для вычисления среднего значения.

Точно так же доказывается и более общая формула для не обязательно непрерывного случая:

$$\mathbb{E}g(\xi_1, \dots, \xi_n) = \int_{\mathbf{R}^n} g(x_1, \dots, x_n) d\mathbb{P}_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n). \quad (2)$$

Но для вычислений она используется редко.

Замечание. Из формулы замены переменной следует тривиальное (?) утверждение: если $\mathbb{P}_\xi = \mathbb{P}_\eta$, то $\mathbb{E}g(\xi) = \mathbb{E}g(\eta)$ для любой измеримой функции g , для которой существует одна из частей равенства.

Замечание. Интеграл Лебега $\int f(x) dm(x)$, где m — мера Лебега, это по-существу интеграл $\int y df \cdot m(y)$, то есть интеграл от тождественной функции по мере $f \cdot m$ — образу меры m относительно отображения f . В теории интегрирования одним из важнейших приемов является замена переменной, которая позволяет упрощать вид интегрируемой функции. Как мы видим, в теории интеграла Лебега мы до предела упрощаем интегрируемую функцию, зато всю сложность переводим в меру.

Приведем нужные объяснения: обычный интеграл Римана является пределом интегральных сумм $\sum_{i=1}^n f(\theta_i)(x_i - x_{i-1})$, где $\theta_i \in [x_{i-1}, x_i]$ — произвольные точки в интервале, используя обозначение меры Лебега эту сумму можно записать в виде $\sum_{i=1}^n f(\theta_i)m([x_{i-1}, x_i])$, где m — мера Лебега, то есть длина соответствующего интервала. Интегральная сумма Лебега имеет вид $\sum_{i=1}^n \eta_i f \cdot m([y_{i-1}, y_i])$, $\eta_i \in [y_{i-1}, y_i]$, а $f \cdot m([y_{i-1}, y_i]) = m(f^{-1}[y_{i-1}, y_i])$. В хорошей ситуации, когда функция f монотонна,

$$\sum_{i=1}^n \eta_i f \cdot m([y_{i-1}, y_i]) = \sum_{i=1}^n f(\theta_i)m([f^{-1}(y_{i-1}), f^{-1}(y_i)]),$$

где $f(\theta_i) = \eta_i$, а в правой части — сумма Римана.

В интеграле Лебега больший произвол интегральных сумм, разбиение области интегрирования не обязательно состоит из интервалов. Поэтому это более общее понятие. Однако вычисление интеграла Лебега производится лишь в том случае, когда его можно записать в виде интеграла Римана. Впрочем, имеется случай, когда интеграл Римана не нужен, это случай кусочно постоянной функции $f \in \{y_1, \dots, y_n\}$ или $f \in \{y_1, \dots, y_n, \dots\}$, в этом случае образ меры m лежит на $\{y_1, \dots, y_n\}$ или $\{y_1, \dots, y_n, \dots\}$, причем $f \cdot m(\{y_i\}) = m(f^{-1}(\{y_i\}))$. Используя фигурные скобки, мы обращаем внимание на то, что y_i здесь не число, а одноточечное множество, прообраз этого множества тоже множество. В этой ситуации вместо интеграла Римана для вычисления мы используем сумму $\sum_{i=1}^n y_i m(f^{-1}(\{y_i\}))$ или сумму ряда $\sum_n y_n m(f^{-1}(\{y_n\}))$.

Точно так же вводится и, когда это возможно, вычисляется среднее значение или математическое ожидание

$$\mathbb{E} \xi = \mathbb{M}\xi = \int_{\Omega} \xi d\mathbb{P}.$$

но в этой ситуации мера \mathbb{P} не мера Лебега, а вероятность, и задана не на отрезках, а на событиях, однако переход к образу вероятности \mathbb{P} — распределению \mathbb{P}_ξ случайной величины ξ позволяет в случае, когда распределение \mathbb{P}_ξ имеет плотность p_ξ , сводить среднее к интегралу Римана, а когда распределение лежит на конечном или счетном множестве, — к конечной сумме или сумме ряда.

Итак, вероятность (и интеграл по ней) — это некая абстракция, а распределение случайной величины мы можем использовать для конкретных вычислений.

4. Представление среднего произведения независимых случайных величин.

Теорема. Пусть ξ, η — независимые случайные величины, $\mathbb{E}\xi$ и $\mathbb{E}\eta$ существуют. Тогда существует и $\mathbb{E}\xi\eta$ и

$$\mathbb{E}\xi\eta = \mathbb{E}\xi\mathbb{E}\eta. \quad (3)$$

Это равенство очевидно для дискретных независимых случайных величин $\xi = \sum_n a_n I_{\{\xi=a_n\}}$ и $\eta = \sum_m b_m I_{\{\eta=b_m\}}$. Проверим:

$$\mathbb{E}\xi = \sum_n a_n \mathbb{P}\{\xi = a_n\},$$

$$\mathbb{E}\eta = \sum_m b_m \mathbb{P}\{\eta = b_m\},$$

$$\mathbb{E}\xi\eta = \sum_{n,m} a_n b_m \mathbb{P}\{\xi = a_n, \eta = b_m\}.$$

Ввиду независимости,

$$\mathbb{P}\{\xi = a_n, \eta = b_m\} = \mathbb{P}\{\xi = a_n\} \mathbb{P}\{\eta = b_m\},$$

поэтому двойной ряд по m и n является произведением однократных рядов. Общий случай как-будто бы легко сводится к дискретному приближению случайными величинами $\xi^{(n)}$ и $\eta^{(n)}$. Легко видеть, что эти случайные величины являются функциями независимых случайных величин ξ и η , поэтому сами являются независимыми, мы знаем, что $\xi^{(n)}$ и $\eta^{(n)}$ равномерное сходятся соответственно к ξ и η . К сожалению, ввиду возможной неограниченности ξ и η , $\xi\eta$ не обязательно является равномерным пределом последовательности $(\xi^{(n)}\eta^{(n)})$. Вместо равномерной сходимости мы воспользуемся сохранением неравенств при интегрировании. Для этого мы разложим

$$\xi = \xi^+ - \xi^-, \eta = \eta^+ - \eta^-,$$

где $\xi^+ = \max\{\xi, 0\}$, $\xi^- = -\min\{\xi, 0\}$, аналогично определяются η^+ η^- , причем опять эти величины независимы как функции независимых величин. В силу аддитивности среднего, нам достаточно проверить (3) лишь для неотрицательных случайных величин ξ и η . В этом случае

$$\begin{aligned} \mathbb{E}\xi^{(n)}\mathbb{E}\eta^{(n)} &= \mathbb{E}\xi^{(n)}\eta^{(n)} \leq \mathbb{E}(\xi\eta) \leq \mathbb{E}\left(\xi^{(n)} + \frac{1}{n}\right)\left(\eta^{(n)} + \frac{1}{n}\right) = \\ &= \mathbb{E}\left(\xi^{(n)} + \frac{1}{n}\right)\mathbb{E}\left(\eta^{(n)} + \frac{1}{n}\right). \end{aligned}$$

Переходя к пределу и учитывая, что

$$\lim_n \mathbb{E}\xi^{(n)} = \lim_n \mathbb{E}\left(\xi^{(n)} + \frac{1}{n}\right) = \mathbb{E}(\xi),$$

$$\lim_n \mathbb{E}\eta^{(n)} = \lim_n \mathbb{E}\left(\eta^{(n)} + \frac{1}{n}\right) = \mathbb{E}(\eta),$$

получаем требуемое равенство.

Упражнение. Если ξ и η имеют совместное невырожденное нормальное распределение и верно (3), то ξ и η независимы.

Замечание. Среднее значение случайной величины – это константа, поэтому $\mathbb{E}[\mathbb{E}\xi] = \mathbb{E}\xi$.

5. Пример использования понятия математического ожидания.

Пусть $\xi_1, \dots, \xi_n, \dots$ — последовательность независимых одинаково распределенных случайных величин. Тогда при очень общих предположениях усредненная сумма $\frac{\xi_1 + \dots + \xi_n}{n}$ сходится к математическому ожиданию случайных величин ξ_n . Это объясняет термин *математическое ожидание*. Соответствующие теоремы *законы больших чисел* будут доказаны в конце курса теории вероятностей. Ниже мы приведем пример из книги Феллера, когда само введение соответствующего понятия приводит к значительной экономии средств.

Пример из книги Феллера. В американской армии кровь новобранцев проверяется на наличие некоторого фактора, вероятность наличия которого примерно равна 0.1. Можно проверять

кровь каждого человека (в этом случае для n человек проводится n анализов), а можно разбить всех новобранцев на группы по k человек. В каждой группе часть отобранный крови сливаются, и если в группе нет людей с данным фактором, то это обстоятельство обнаруживается в результате одного анализа. Если же такие люди в группе имеются, то проводится $1 + k$ анализов. Как нужно разбивать на группы, чтобы минимизировать математическое ожидание общего числа анализов? Решение заключается в нахождении минимума последовательности

$$\mathbb{E}\{\text{среднее число анализов для одного человека}\} = \frac{1 \times 0.9^k + (k+1)(1-0.9^k)}{k}.$$

Ответ: $k = 4$.

6. Пример из теории игр. Теорема фон Неймана о минимаксе.

Пример. Играют два игрока, у каждого из которых имеется r возможных стратегий игры. Если 1-й игрок использует i -ю стратегию, а 2-й игрок использует против этой стратегии свою j -ю стратегию, то выигрыш 1-го игрока составит x_{ij} (это число может быть как положительным, так и отрицательным). Оба игрока знают матрицу $[x_{ij}]$. Как должен играть 1-й игрок, чтобы получить максимальный выигрыш при любой стратегии противника? Заметим, что пока в условиях нет никаких вероятностей и все детерминировано.

Давайте рассмотрим конкретную матрицу:

$$[x_{ij}] = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 3 \\ 4 & 3 & 2 \end{bmatrix}.$$

Как видно из матрицы, при выборе любой стратегии при правильной игре противника выигрыш 1-го игрока не больше 2. Но давайте сделаем выбор стратегии случайнym: каждую из стратегий мы выбираем с вероятностью $1/3$, и противник не знает, какую стратегию на каждом шаге мы выбрали. Тогда выигрыш 1-го игрока становится случайной величиной: при выборе 2-м игроком его 1-й стратегии математическое ожидание выигрыша 1-го игрока равно $\frac{1+3+4}{3} = \frac{8}{3}$, при выборе 2-м игроком его 2-й стратегии математическое ожидание выигрыша 1-го игрока равно $\frac{2+2+3}{3} = \frac{7}{3}$, при выборе 2-м игроком его 1-й стратегии математическое ожидание выигрыша 1-го игрока равно $\frac{3+3+1}{3} = \frac{7}{3}$. Итак, вероятностная стратегия (в теории игр такая стратегия называется *смешанной*) даже при первом попавшемся подборе вероятностей оказывается более выгодной. Наименьший результат будет равен $7/3$. Еще лучше будет результат 1 игрока, если он выберет 1-ю стратегию с вероятностью $p_1 = 1/3$, 2-ю стратегию с вероятностью $p_2 = 4/15$, 3-ю стратегию с вероятностью $p_3 = 2/5$. Однако, 2-й игрок также может выбрать смешанную стратегию, состоящую из вероятности q_1 1-й стратегии, q_2 — 2-й стратегии, q_3 — 3-й стратегии. Тогда общий выигрыш равен $\sum_{i,j=1}^3 x_{ij} p_i q_j$. В более общей ситуации задача 1 игрока — выбрать такую смешанную стратегию $\{p_i\}$, чтобы число $\min_{\sum q_j=1} \sum_{i,j} x_{ij} p_i q_j$ было бы максимально, а задача 2 игрока — выбрать такую смешанную стратегию $\{q_j\}$, чтобы число $\max_{\sum p_i=1} \sum_{i,j} x_{ij} p_i q_j$ было бы минимально. Таким образом, при правильной игре 1-го игрока он обеспечит себе выигрыш не меньше $\max_{\sum p_i=1} \min_{\sum q_j=1} \sum_{i,j} x_{ij} p_i q_j$, а при правильной игре 2-го игрока он обеспечит себе проигрыш не больше $\min_{\sum q_j=1} \max_{\sum p_i=1} \sum_{i,j} x_{ij} p_i q_j$.

Замечательная теорема фон Неймана утверждает, что

$$\max_{\sum p_i=1} \min_{\sum q_j=1} \sum_{i,j} x_{ij} p_i q_j = \min_{\sum q_j=1} \max_{\sum p_i=1} \sum_{i,j} x_{ij} p_i q_j.$$

7. Дисперсия.

Дисперсией случайной величины ξ называется

$$\mathbb{D}\xi = \mathbb{E}(\xi - \mathbb{E}\xi)^2.$$

Для дисперсии имеется и вторая формула

$$\mathbb{D}\xi = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2.$$

Проверка очевидна:

$$\mathbb{E}(\xi - \mathbb{E}\xi)^2 = \mathbb{E}\xi^2 - 2\mathbb{E}(\xi\mathbb{E}\xi) + \mathbb{E}(\mathbb{E}\xi)^2 = \mathbb{E}\xi^2 - 2\mathbb{E}\xi\mathbb{E}\xi + \mathbb{E}\xi\mathbb{E}\xi.$$

Из определения следует: *дисперсия — мера отклонения случайной величины от своего среднего*.

8. Свойства дисперсии (предполагается, что дисперсия существует).

1. $\mathbb{D}\xi \geq 0$.
2. $\mathbb{D}(\xi + Const) = \mathbb{D}\xi$.
3. $\mathbb{D}(a\xi) = a^2\mathbb{D}\xi$.
4. Если ξ и η — независимые случайные величины, то

$$\mathbb{D}(\xi + \eta) = \mathbb{D}\xi + \mathbb{D}\eta.$$

Проверка очевидна.

$$\begin{aligned} & \mathbb{E}([\xi - \mathbb{E}\xi] + [\eta - \mathbb{E}\eta])^2 = \\ & \mathbb{E}(\xi - \mathbb{E}\xi)^2 + \mathbb{E}(\eta - \mathbb{E}\eta)^2 + 2\mathbb{E}([\xi - \mathbb{E}\xi][\eta - \mathbb{E}\eta]). \end{aligned}$$

Но

$$\mathbb{E}([\xi - \mathbb{E}\xi][\eta - \mathbb{E}\eta]) = \mathbb{E}[\xi - \mathbb{E}\xi]\mathbb{E}[\eta - \mathbb{E}\eta] = 0 \cdot 0.$$

Разумеется, не всегда дисперсия суммы равна сумме дисперсий, например,

$$\mathbb{D}(\xi + \xi) = \mathbb{D}(2\xi) = 4\mathbb{D}\xi \neq \mathbb{D}\xi + \mathbb{D}\xi.$$

Упражнения. 1. Если $\mathbb{P}_\xi = \mathbb{P}_\eta$, то $\mathbb{E}\xi = \mathbb{E}\eta$, $\mathbb{D}\xi = \mathbb{D}\eta$. Обратное неверно — придумайте пример.

2. Если ξ_1, \dots, ξ_n — независимые одинаково распределенные случайные величины со средним и дисперсией, то

$$\mathbb{E}\left(\frac{\xi_1 + \dots + \xi_n}{n}\right) = \mathbb{E}\xi_1 = \dots = \mathbb{E}\xi_n,$$

$$\mathbb{D}\left(\frac{\xi_1 + \dots + \xi_n}{\sqrt{n}}\right) = \mathbb{D}\xi_1 = \dots = \mathbb{D}\xi_n.$$

Будут ли верны эти равенства, если случайные величины ξ_1, \dots, ξ_n попарно независимы, т. е. ξ_i, ξ_j независимы для любых пар $i \neq j$?

9. Моменты.

n-м моментом случайной величины называется

$$\alpha_n(\xi) = \mathbb{E}\xi^n.$$

Замена ξ случайной величиной $\xi - \mathbb{E}(\xi)$ делает среднее значение равным нулю и называется *центрированием* случайной величины.

n-м центральным моментом случайной величины называется

$$\mu_n(\xi) = \mathbb{E}[\xi - \mathbb{E}\xi]^n.$$

Среднее значение — это первый момент. *Дисперсия* — второй центральный момент. Очевидно, что n-й момент при умножении случайной величины на a умножается на a^n , кроме

того, n -й центральный момент не меняется при прибавлении к случайной величине константы.

10. Другие характеристики.

Медиана распределения \mathbb{P} — это такое число d , что

$$\mathbb{P}(-\infty, d] \geq \frac{1}{2}, \mathbb{P}[d, \infty) \geq \frac{1}{2}.$$

Медиану распределения может быть несколько.

В математической статистике постоянно используется квантиль случайной величины: *квантиль порядка α* случайной величины ξ это любое такое число K_α , что

$$\mathbb{P}\{\xi < K_\alpha \leq \alpha, \mathbb{P}\{\xi \leq K_\alpha \geq \alpha\}.$$

Впрочем, используются также таблицы чисел $K_{1-\alpha}$.

Модой дискретного распределения называется точка, вероятность которой — наибольшая. *Модой* непрерывного распределения называется точка, в которой плотность — наибольшая. Если у распределения мода единственна, оно называется *унимодальным*.

11. Среднее и дисперсия конкретных распределений.

Вычисление среднего и дисперсии можно проводить как непосредственно, в виде суммы или интеграла, так и с использованием свойств среднего и дисперсии, в частности, для среднего суммы случайных величин и для дисперсии суммы независимых случайных величин

Распределение Бернулли с параметром p .

$$\mathbb{E}\xi = \int xd\mathbb{P}_\xi(x) = 1 \cdot \mathbb{P}\{\xi = 1\} + 0 \cdot \mathbb{P}\{\xi = 0\} = p.$$

$$\mathbb{D} = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2 = p - p^2.$$

Биномиальное распределение с символом $B(n, p)$.

$$\begin{aligned} \mathbb{E} &= \int xd\mathbb{P}(x) = \sum_{k=0}^n k\mathbb{P}\{k\} = \sum_{k=0}^n kC_n^k p^k (1-p)^{n-k} = \\ &= \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = np \sum_{k=1}^n C_n^{k-1} p^{k-1} (1-p)^{n-k} = np. \end{aligned}$$

Аналогичные выкладки позволяют подсчитать и второй момент. Однако наш ответ слишком прост, нельзя ли его получить проще. Вспомним, что биномиальное распределение — это распределение числа гербов в n испытаниях. Эту случайную величину ξ мы можем представить как сумму:

$$\xi = \delta_1 + \dots + \delta_n,$$

где δ_i — число гербов в i испытании имеет среднее p и дисперсию $p(1-p)$. Используем свойства среднего и дисперсии (для независимых слагаемых), получаем, $\mathbb{E} = np$, $\mathbb{D} = np(1-p)$.

Упражнение. Найдите моду биномиального распределения. Укажите алгоритм нахождения для любого p .

Распределение Пуассона с параметром λ .

$$\begin{aligned} \mathbb{E}\xi &= \int xd\mathbb{P}_\xi(x) = \sum_{k=0}^{\infty} k\mathbb{P}\{\xi = k\} = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \\ &= \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \lambda e^\lambda e^{-\lambda} = \lambda. \end{aligned}$$

$$\begin{aligned}\mathbb{E}\xi^2 &= \sum_{k=0}^{\infty} k^2 \mathbb{P}\{\xi = k\} = \sum_{k=0}^{\infty} \{k(k-1) + k\} \frac{\lambda^k}{k!} e^{-\lambda} = \\ &= \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} e^{-\lambda} + \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \lambda^2 + \lambda. \\ \mathbb{D}\xi &= \lambda^2 + \lambda - \lambda^2 = \lambda.\end{aligned}$$

Равномерное распределение на отрезке $[a, b]$.

$$\begin{aligned}\mathbb{E}\xi &= \int_a^b \frac{1}{b-a} x dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}. \\ \mathbb{E}\xi^2 &= \int_a^b \frac{1}{b-a} x^2 dx = \frac{b^3 - a^3}{3(b-a)}. \\ \mathbb{D}\xi &= \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.\end{aligned}$$

12. Характеристики нормального распределения.

Проще всего все подсчеты сделать для стандартного нормального распределения $N(0, 1)$, общий случай свести к нему.

$$\mathbb{E}\xi = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x e^{-x^2/2} dx = 0$$

как интеграл от нечетной функции. Второй момент интегрируем по частям.

$$\begin{aligned}\mathbb{D}\xi = \mathbb{E}\xi^2 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-x^2/2} dx = - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x de^{-x^2/2} = \\ &= -\frac{1}{\sqrt{2\pi}} x de^{-x^2/2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0 + 1 = 1.\end{aligned}$$

Мода и медиана ξ равны 0. Подсчитаем высшие моменты. Для $n = 2k + 1$

$$\mathbb{E}\xi^{2k+1} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x^{2k+1} e^{-x^2/2} dx = 0.$$

Для $n = 2k$ интегрируем по частям:

$$\begin{aligned}\alpha_{2k}(\xi) = \mathbb{E}\xi^{2k} &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x^{2k} e^{-x^2/2} dx = \\ &= -\frac{1}{\sqrt{2\pi}} x^{2k-1} de^{-x^2/2} \Big|_{-\infty}^{\infty} + (2k-1) \int_{-\infty}^{\infty} x^{2k-2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \\ &\quad \alpha_{2k-2}(\xi) = \dots = (2k-1)!!.\end{aligned}$$

В случае $\eta = \sigma\xi + m$ (т. е. в случае $N(m, \sigma)$) имеем:

$$\mathbb{E}\eta = m, \mathbb{D}\eta = \sigma^2, \mu_{2k}(\eta) = (2k-1)!!\sigma^{2k}.$$

Упражнение. Разложите по биному Ньютона $\eta^n = (\sigma\xi + m)^n$. Сложив средние отдельных слагаемых, выведите формулу для $\alpha_n(\eta)$.

13. Среднее и дисперсия распределения χ_n^2 .

Для вычислений надо использовать не функцию плотности распределения χ_n^2 , а определение этого распределения χ_n^2 : $\chi_n^2 = \xi_1^2 + \dots + \xi_n^2$, где ξ_i — независимые случайные величины со стандартным нормальным распределением. Поэтому

$$\mathbb{E}\chi_n^2 = \mathbb{E}\xi_1^2 + \dots + \mathbb{E}\xi_n^2 = n \cdot 1.$$

$$\mathbb{D}\chi_n^2 = \mathbb{D}\xi_1^2 + \dots + \mathbb{D}\xi_n^2 = n \cdot \left(\mathbb{E}\xi_1^4 - (\mathbb{E}\xi_1^2)^2 \right) = n(3-1) = 2n.$$

14. Случай распределения Коши.

Для распределение Коши мы имеем следующий интеграл для вычисления среднего:

$$\int_{-\infty}^{\infty} x \frac{1}{\pi(1+x^2)} dx.$$

Если бы интеграл сходился, то он был бы равен нулю, но он не сходится, среднее не существует.

15. Смешанные моменты

Для n случайных величин ξ_1, \dots, ξ_n обычно записывают вектор средних и матрицу вторых центральных моментов (если существуют) $[\mu_{i,j}]$, где

$$\mu_{i,j} = \mathbb{E}[\xi_i - \mathbb{E}(\xi_i)][\xi_j - \mathbb{E}(\xi_j)]$$

Матрица $[\mu_{i,j}]$ неотрицательно определена. Действительно, для всех вещественных x_i

$$\sum_{i,j} \mu_{i,j} x_i x_j = \mathbb{E} \left[\sum_i (\xi_i - \mathbb{E}\xi_i) x_i \right]^2 \geq 0.$$

Упражнения. а) Используя неравенство Коши — Буняковского для средних значений докажите, что из существования дисперсий ξ и η следует и существование у них центрального смешанного момента.

б) Обобщите свойство аддитивности для дисперсии независимых случайных величин: покажите, что из независимости случайных векторов $(\xi_1, \eta_1), (\xi_2, \eta_2)$ и существования вторых смешанных моментов следует равенство

$$\mathbb{E}[(\xi_1 + \xi_2 - \mathbb{E}(\xi_1 + \xi_2))(\eta_1 + \eta_2 - \mathbb{E}(\eta_1 + \eta_2))] = \mathbb{E}[(\xi_1 - \mathbb{E}\xi_1)(\eta_1 - \mathbb{E}\eta_1)] + \mathbb{E}[(\xi_2 - \mathbb{E}\xi_2)(\eta_2 - \mathbb{E}\eta_2)].$$

Можно ли доказать равенство

$$\mathbb{E}[(\xi_1 + \xi_2 - \mathbb{E}(\xi_1 + \xi_2))(\eta_1 + \eta_2 - \mathbb{E}(\eta_1 + \eta_2))] = \mathbb{E}[(\xi_1 - \mathbb{E}\xi_1)(\eta_2 - \mathbb{E}\eta_2)] + \mathbb{E}[(\xi_2 - \mathbb{E}\xi_2)(\eta_1 - \mathbb{E}\eta_1)]?$$

16. Коэффициент корреляции двух случайных величин ξ и η . Если случайные величины ξ и η независимы, то смешанный центральный момент

$$\mu(\xi, \eta) = \mathbb{E}[\xi - \mathbb{E}\xi][\eta - \mathbb{E}\eta]$$

равен 0. Но $\mu(\xi, \eta)$ не может характеризовать степень зависимости случайных величин, так как сам зависит от внешних факторов (например, единиц измерения, изменение которых может во много раз изменить этот коэффициент). Этого недостатка лишен коэффициент корреляции:

$$\rho(\xi, \eta) = \frac{\mathbb{E}[\xi - \mathbb{E}\xi][\eta - \mathbb{E}\eta]}{\sqrt{\mathbb{D}(\xi)\mathbb{D}(\eta)}}.$$

Свойства коэффициента корреляции

1. $\rho(a\xi + b, c\eta + d) = \text{sgn}(ac)\rho(\xi, \eta)$ для всех вещественных a, b, c, d .
2. $|\rho(\xi, \eta)| \leq 1$.
3. Если ξ и η независимы, то $\rho(\xi, \eta) = 0$, $|\rho(\xi, \eta)| = 1$ тогда и только тогда, когда ξ и η линейно зависимы, т.е. $a\xi + b\eta + c = 0$ для некоторых вещественных a, b, c .

Чтобы удобнее было доказать 1, мы введем операцию

$$\xi \rightarrow \xi^* = \frac{\xi - \mathbb{E}\xi}{\sqrt{\mathbb{D}(\xi)}}$$

центрирования и нормирования случайной величины ξ . Легко проверяется, что

$$\rho(\xi, \eta) = \mathbb{E}\xi^*\eta^*, (a\xi + b)^* = \text{sgn } a\xi^*.$$

Это влечет 1. Свойство 2 следует из неравенства Коши — Буняковского

$$|\mathbb{E}\xi\eta| \leq \sqrt{\mathbb{E}|\xi|^2\mathbb{E}|\eta|^2}.$$

Первую часть 3 мы уже отметили. Чтобы доказать вторую часть, нужно вспомнить — когда в неравенстве Коши — Буняковского достигается равенство. Тогда и только тогда, когда ξ и η пропорциональны. В наших терминах это означает $\xi^* = C\eta^*$, это влечет утверждение в 3. Обратно, из введенного соотношения между ξ и η следует $\text{sgn } a\xi^* = -\text{sgn } b\eta^*$.

Равенство нулю коэффициента корреляции $\rho(\xi, \eta)$ не влечет независимость ξ и η .

Пример. $\Omega = \{1, 2, 3, 4\}$ с классической вероятностью, $\xi(1) = 1, \xi(2) = -1, \xi(3) = \xi(4) = 0, \eta(1) = \eta(2) = 0, \eta(3) = 1, \eta(4) = -1$. Тогда средние ξ и η равны 0, $\xi\eta = 0$ тождественно. Разумеется, случайные величины ξ и η зависимы.

Предложение. *Если ξ и η имеют совместное нормальное распределение, $\rho(\xi, \eta) = 0$, то ξ и η независимы.*

Мы будем считать для простоты ξ и η центрированными, тогда их совместная функция плотности имеет вид

$$p_{\xi, \eta}(x, y) = \frac{|A|}{2\pi} e^{-\frac{1}{2}(a_{11}x^2 + 2a_{12}xy + a_{22}y^2)}.$$

Чтобы доказать независимость, нам нужно показать, что $a_{12} \neq 0$. Это можно сделать и непосредственно — вычислением $\rho(\xi, \eta) = 0$, а также анализом интеграла, который мы сейчас продемонстрируем. Числитель в ρ имеет вид

$$\int \int xy \frac{|A|}{2\pi} e^{-\frac{1}{2}(a_{11}x^2 + 2a_{12}xy + a_{22}y^2)} dx dy.$$

Сравним этот интеграл с равным нулю интегралом

$$\int \int xy \frac{|A|}{2\pi} e^{-\frac{1}{2}(a_{11}x^2 + a_{22}y^2)} dx dy.$$

Допустим, что $a_{12} > 0$. Разобьем оба интеграла на две области — $\{xy > 0\}$ и $\{xy < 0\}$. Непосредственно проверяется, что на каждой из областей верхний интеграл строго меньше нижнего. Поэтому оба интеграла не могут быть одновременно равны 0. Случай $a_{12} < 0$ рассматривается аналогично.

Пример. Пусть $\mathbb{P}_{\xi, \eta}$ имеет плотность, равную $1/\pi$ на круге $\{(x, y) : x^2 + y^2 \leq 1\}$ и 0 вне круга. Докажите, что $\rho(\xi, \eta) = 0$, тем не менее, случайные величины ξ и η зависимы. (Указание. Для доказательства зависимости придумайте такие числовые множества A и B , что

$$\mathbb{P}\{\xi \in A\} = \mathbb{P}_\xi(A) \neq 0, \mathbb{P}\{\eta \in B\} = \mathbb{P}_\eta(B) \neq 0,$$

$$\mathbb{P}(\{\xi \in A\} \cap \{\eta \in B\}) = \mathbb{P}_{\xi, \eta}(A \times B) = 0.)$$

А что будет, если

$$p_{\xi,\eta}(x,y) = \begin{cases} a, & \text{если } x^2 + y^2 \leq 1, \\ b & \text{если } |x| \leq 1, |y| \leq 1, x^2 + y^2 > 1, \\ 0 & \text{если } |x| > 1 \text{ или } |y| > 1? \end{cases}$$

Как связаны числа a и b ?

Упражнение 14. Случайные величины ξ и η независимы, $\mathbb{D}(\xi) = 1$,

$$\mathbb{E}(\xi) = \mathbb{D}(\xi) = \mathbb{E}(\eta) = \mathbb{D}(\eta) = 1.$$

Найти $\mathbb{E}(\xi - \eta)$, $\mathbb{D}(\xi - \eta)$.

Упражнение 15. Случайные величины ξ и η имеют совместное нормальное распределение, $\mathbb{D}(\xi + \eta) = \mathbb{D}(\xi) + \mathbb{D}(\eta)$. Доказать, что ξ и η независимы.

Вопрос. Придумайте пример двух случайных величин с разными распределениями и одинаковыми средними и дисперсиями.

Вопрос. В каком классе распределений среднее и дисперсия совпадают у всех распределений из класса?

§8. Задача регрессии

1. Функция регрессия.

Мы будем рассматривать задачу приближения одной случайной величины η функцией $f(\xi_1, \xi_2, \dots, \xi_n)$ других случайных величин $\xi_1, \xi_2, \dots, \xi_n$. А именно, мы будем минимизировать число $\mathbb{E}(\eta - f(\xi_1, \xi_2, \dots, \xi_n))^2$ (см. объяснение задачи также в лекции о сходимости).

Сначала рассмотрим простейший случай, когда ξ одно и является константой. В этом случае решение очень просто. Если $\xi = c$, то функция f определена лишь на множестве, состоящем из одной точки $\{c\}$. Оказывается, что минимум $\mathbb{E}(\eta - f(c))^2$ достигается при $f(c) = \mathbb{E}\eta$. Действительно, предположим $f(c) = d \neq \mathbb{E}\eta$. Тогда

$$\mathbb{E}(\eta - d)^2 = \mathbb{E}(\eta - \mathbb{E}\eta + \mathbb{E}\eta - d)^2 = \mathbb{E}(\eta - \mathbb{E}\eta)^2 + \mathbb{E}(\mathbb{E}\eta - d)^2 + \mathbb{E}(\eta - \mathbb{E}\eta)(\mathbb{E}\eta - d).$$

Заметим, что $\mathbb{E}\eta - d$ является константой, поэтому

$$\mathbb{E}(\eta - \mathbb{E}\eta)(\mathbb{E}\eta - d) = (\mathbb{E}\eta - d)\mathbb{E}(\eta - \mathbb{E}\eta) = 0.$$

Последнее равенство основано на том, что $\mathbb{E}(\mathbb{E}\eta) = \mathbb{E}\eta$. Это верно лишь при интегрировании по вероятностной мере,

$$\mathbb{E}c = \int_{\Omega} c = c\mathbb{P}(\Omega) = c.$$

Далее мы рассмотрим общую ситуацию дискретной случайной величины ξ :

$$\xi = \sum_n a_n I_{\{\xi=a_n\}}.$$

Минимизируемую величину $\mathbb{E}(\eta - f(\xi))^2$ нагляднее записать в виде интеграла

$$\mathbb{E}(\eta - f(\xi))^2 = \int_{\Omega} (\eta - f(\xi))^2 d\mathbb{P} =$$

$$\sum_n \int_{\{\xi=a_n\}} \left(\eta - f \left(\sum_n a_n I_{\{\xi=a_n\}} \right) \right)^2 d\mathbb{P}.$$

Как и выше, мы видим, что существенными являются лишь значения функции f на числах a_n , то есть f задается набором чисел $b_n = f(a_n)$ и это значения случайной величины $f(\xi)$ на событиях $\{\xi = a_n\}$. Имеем

$$\mathbb{E}(\eta - f(\xi))^2 = \int_{\Omega} (\eta - f(\xi))^2 d\mathbb{P} = \sum_n \int_{\{\xi=a_n\}} (\eta - b_n)^2 d\mathbb{P}.$$

Сумма окажется минимальной, если каждое слагаемое минимально. Таким образом, нам нужно минимизировать

$$\int_{\{\xi=a_n\}} (\eta - b_n)^2 d\mathbb{P}. \quad (1)$$

В результате, мы опять пришли к случаю константы, но интегрирование ведется по событию $\{\xi = a_n\}$, а мера \mathbb{P} этого события не равна 1 и ее сужение не является вероятностью. Чтобы использовать результат для константы, мы должны разделить сужение меры \mathbb{P} на множество $\{\xi = a_n\}$ на вероятность $\mathbb{P}(\{\xi = a_n\})$. Новая мера $\mathbb{P}_n(B) = \frac{\mathbb{P}(B|\{\xi = a_n\})}{\mathbb{P}(\{\xi = a_n\})}$ является вероятностью. Поэтому минимум (1) достигается при

$$b_n = \int_{\{\xi=a_n\}} \eta d\mathbb{P}_n. \quad (1).$$

Заметим, что мы можем записать новую вероятность как условную:

$$\mathbb{P}_n(B) = \mathbb{P}(B|\{\xi = a_n\}).$$

2. Случай совместного распределения непрерывного типа.

В этом случае имеется совместная функция плотности $p_{\xi,\eta}$ случайных величин ξ, η , зависящая от двух вещественных аргументов x, y . Минимизируемое число $\mathbb{E}(\eta - f(\xi))^2$ выражается через $p_{\xi,\eta}$ по формуле замены переменной следующим образом:

$$\mathbb{E}(\eta - f(\xi))^2 = \int_{\mathbf{R} \times \mathbf{R}} (y - f(x))^2 p_{\xi,\eta}(x, y) dx dy.$$

Как и в дискретном случае, мы хотим свести задачу минимизации к случаю константы, в данном случае это означает фиксацию переменной x , что возможно переходом к повторному интегралу:

$$\int_{\mathbf{R} \times \mathbf{R}} (y - f(x))^2 p_{\xi,\eta}(x, y) dx dy = \int_{\mathbf{R}} dx \int_{\mathbf{R}} (y - f(x))^2 p_{\xi,\eta}(x, y) dy.$$

Чтобы минимизировать этот интеграл, достаточно минимизировать внутренний интеграл при любом фиксированном x . Но случай константы здесь не подходит, так как мера $p_{\xi,\eta}(x, y) dy$, по которой мы интегрируем функцию, не является вероятностной. Давайте сделаем ее вероятностной, разделив при данном x функцию $p_{\xi,\eta}(x, y)$ на интеграл по этой функции. Заодно мы введем условную плотность

$$p_{\eta}(y|\xi = x) = \frac{p_{\xi,\eta}(x, y)}{\int_{\mathbf{R}} p_{\xi,\eta}(x, y) dy}.$$

Теперь (применив случай постоянной к элементарным исходам y , где $\eta(y) = y$, $\xi(y) = x$), мы получаем нужное решение:

$$f(x) = \frac{\int_{\mathbf{R}} y p_{\xi,\eta}(x, y) dy}{\int_{\mathbf{R}} p_{\xi,\eta}(x, y) dy} = \int_{\mathbf{R}} y p_{\eta}(y|\xi = x) dy.$$

3. Случай совместного нормального распределения.

В нормальном случае

$$p_{\xi,\eta}(x, y) = \frac{|A|^{1/2}}{2\pi} e^{-\frac{1}{2}(a_{11}(x-m_x)^2 + 2a_{12}(x-m_x)(y-m_y) + a_{22}(y-m_y)^2)},$$

где $A = [a_{ij}]$, $|A|$ — определитель A . Имеем:

$$f(x) = \frac{\int y \frac{|A|^{1/2}}{2\pi} e^{-\frac{1}{2}(a_{11}(x-m_x)^2 + 2a_{12}(x-m_x)(y-m_y) + a_{22}(y-m_y)^2)} dy}{\int \frac{|A|^{1/2}}{2\pi} e^{-\frac{1}{2}(a_{11}(x-m_x)^2 + 2a_{12}(x-m_x)(y-m_y) + a_{22}(y-m_y)^2)} dy}.$$

Мы сокращаем в числителе и знаменателе множители, не зависящие от y :

$$f(x) = \frac{\int y e^{-\frac{1}{2}(2a_{12}(x-m_x)(y-m_y) + a_{22}(y-m_y)^2)} dy}{\int e^{-\frac{1}{2}(2a_{12}(x-m_x)(y-m_y) + a_{22}(y-m_y)^2)} dy}.$$

Далее мы делаем показатель экспоненты полным квадратом и, наоборот, добавляем слагаемые, зависящие от x . Получаем:

$$f(x) = \frac{\int y e^{-\frac{1}{2}a_{22}\left(\frac{a_{12}}{a_{22}}(x-m_x) + (y-m_y)\right)^2} dy}{\int e^{-\frac{1}{2}a_{22}\left(\frac{a_{12}}{a_{22}}(x-m_x) + (y-m_y)\right)^2} dy}.$$

Далее представляем числитель как сумму двух интегралов, первый из которых после замены переменной оказывается интегралом от нечетной функции, и поэтому равен нулю, а второй после выноса константы равен знаменателю.

$$\frac{\int \left(\frac{a_{12}}{a_{22}}(x-m_x) + (y-m_y)\right) e^{-\frac{1}{2}a_{22}\left(\frac{a_{12}}{a_{22}}(x-m_x) + (y-m_y)\right)^2} dy}{\int e^{-\frac{1}{2}a_{22}\left(\frac{a_{12}}{a_{22}}(x-m_x) + (y-m_y)\right)^2} dy} + \\ \frac{\int \left(-\frac{a_{12}}{a_{22}}(x-m_x) + m_y\right) e^{-\frac{1}{2}a_{22}\left(\frac{a_{12}}{a_{22}}(x-m_x) + (y-m_y)\right)^2} dy}{\int e^{-\frac{1}{2}a_{22}\left(\frac{a_{12}}{a_{22}}(x-m_x) + (y-m_y)\right)^2} dy}.$$

Итак,

$$f(x) = -\frac{a_{12}}{a_{22}}(x-m_x) + m_y.$$

4. Обсуждение.

По-видимому, о применении функции регрессии говорилось в учебнике по истории для 4 класса в пятидесятые годы двадцатого века. В нем рассказывалось, что при бомбеках Москвы немецкие летчики применяли противозенитный маневр, а советские математики придумали метод организации стрельбы, учитывающий этот маневр. Скорее всего, речь шла о вычислении функции $f(\xi_1, \xi_2, \xi_3)$ или $f(\xi_1, \xi_2)$, наилучшим образом приближающее положение η в четвертый или третий момент времени.

5. Метод наименьших квадратов

Пусть задано несколько наблюдений $(x_1^{(i)}, \dots, x_s^{(i)}; y^{(i)})$, $i \leq n$. Задача состоит в приближении переменной y некоторой функцией $f(x_1, \dots, x_s)$. Решение строится методом Гаусса наименьших квадратов, приписывая всем наблюдениям одинаковую меру $1/n$ (результат будет тот же, если это будет считая мера) мы приходим к задаче минимизации суммы квадратов

$$S_n = \sum_{i=1}^n \left(f(x_1^{(i)}, \dots, x_s^{(i)}) - y^{(i)} \right)^2. \quad (*)$$

К сожалению, использованная в п. 1-3 постановка задачи с произвольной функцией f оказывается неразумной, так как векторы $(x_1^{(i)}, \dots, x_s^{(i)})$, как правило, не повторяются, и поэтому ничего лучше бессмысленного решения $f(x_1^{(i)}, \dots, x_s^{(i)}) = y^{(i)}$ мы не получим. Поэтому f нужно выбирать из некоторого класса функций. Расширение класса увеличивает точности оценки, но уменьшает доверие к этой точности. Если мы считаем, что совместное распределение наблюдений нормально, то, согласно пункту 4 имеет смысл рассмотреть самый простой вид функции $f(x_1, \dots, x_s) = a_1 x_1 + \dots + a_s x_s + b$, и все сводится к дифференцированию $(*)$ по переменным a_i и b . В более сложной ситуации есть смысл попробовать в качестве f полином второго порядка. Если мы уверены в периодической зависимости y , надо попробовать тригонометрический полином. В каждой из этой ситуации нужно дифференцировать по неизвестным параметрам, а производные приравнивать нулю.

6. Обобщение условного математического ожидания на общий случай.

Итак, мы увидели что функция регрессии имеет следующие свойства: пусть \mathfrak{B} — σ -подалгебра σ -алгебры \mathfrak{A} , на которой задана случайные величины ξ и η . Известно, что \mathfrak{B} — наименьшая σ -алгебра, относительно которой измерима случайная величина ξ . Тогда функция регрессии $f(\xi)$ будет иметь следующие свойства:

- i) $\int_B f(\xi) d\mathbb{P} = \int_B \eta d\mathbb{P}$ для любого множества $B \in \mathfrak{B}$.
- ii) $f(\xi)$ измерима относительно \mathfrak{B} .

Наша задача — доказать существование такой случайной величины, которую мы (в нашей ситуации) будем называть условным математическим ожиданием η относительно \mathfrak{B} и обозначать $\mathbb{E}(\eta|\mathfrak{B})$.

Мы применяем для этого теорему Радона–Никодима, которая утверждает, что любой заряд ν , непрерывный относительно конечной меры μ (то есть $\mu(A) = 0$ влечет $\nu(A) = 0$), представим в виде неопределенного интеграла $A \rightarrow \int_A \frac{d\nu}{d\mu} d\mu$ от некоторой измеримой функции $\frac{d\nu}{d\mu}$

(которая называется производной Радона–Никодима), причем функция $\frac{d\nu}{d\mu}$ определена единственным образом с точностью до множества μ -меры 0. В нашем случае в качестве заряда мы берем неопределенный интеграл $A \rightarrow \int_A \eta d\mathbb{P}$, но считаем его определенным не на большой σ -алгебре \mathfrak{A} , а на ее σ -подалгебре \mathfrak{B} . Эта функция множеств счетно-аддитивна и равна нулю на множестве нулевой меры, но мерой, вообще говоря, не является, так как может принимать и отрицательные значения. Итак, производная Радона–Никодима является условным математическим ожиданием.

Для полноты изложения мы обязаны доказать теорему Радона–Никодима.

Наиболее трудным моментом доказательства является доказательство следующего утверждения.

Теорема Хана. Пусть ν — ограниченный заряд на σ -алгебре \mathfrak{A} на Ω . Тогда существует такое $B \in \mathfrak{A}$, что $\nu(B) = \sup\{\nu(C) : C \in \mathfrak{A}\}$.

Доказательство. Очевидно, что существует последовательность измеримых (A_n) такая, что $\nu(A_n) \rightarrow \sup \nu$ (по определению супремума числового множества). С конечным набором множеств $(A_i)_{i \leq n}$ мы свяжем разбиение Ω на 2^n попарно непересекающихся множеств $C_j(n)$, каждое из которых представимо в виде $\bigcap_i A_i^*$, где каждое A_i^* равно либо A_i , либо его дополнению. Разумеется, некоторые $C_j(n)$ могут оказаться и пустыми. Определим теперь множество

B_n — объединение всех $C_j(n)$ положительной меры ν . Покажем, что можно принять в качестве B множество

$$B = \bigcap_m \bigcup_{m \leq n < \infty} B_n.$$

Для доказательства отметим сначала, что ввиду свойства непрерывности заряда, которая доказывается так же, как и непрерывность вероятностной меры (см. Предложение 4.11), имеет место сходимость

$$\nu \left(\bigcup_{m \leq n < \infty} B_n \right) \rightarrow \nu(B).$$

Поэтому достаточно установить два неравенства:

- i) $\nu(B_n) \geq \nu(A_n)$ для всех n ,
- ii) $\nu \left(\bigcup_{m \leq n < \infty} B_n \right) \geq \nu(B_m)$ для всех m .

И то и другое неравенство доказывается не выкладками, а логическими рассуждениями. Неравенство i) следует из того, что как событие A_n , так и событие B_n являются объединениями непересекающихся множеств $C_j(n)$, причем в B_n входят все слагаемые A_n положительного ν -заряда, убраны все слагаемые A_n отрицательного ν -заряда и добавлены все слагаемые дополнения A_n положительного ν -заряда. Точно так же для всех $s \geq m$ устанавливается неравенство

$$\nu \left(\bigcup_{m \leq n \leq s} B_n \right) \leq \nu \left(\bigcup_{m \leq n \leq s+1} B_n \right). \quad (*)$$

Действительно, множество в правой части (*) получается из левой части добавлением некоторых множеств $C_j(s+1)$ положительного ν -заряда. Из (*) немедленно следует ii).

Разложение Хана. Если ν — ограниченный заряд на σ -алгебре подмножеств Ω , то

$$\Omega = \Omega^+ + \Omega^-,$$

где $\nu(A) \geq 0$ для любого подмножества $A \subset \Omega^+$ и $\nu(A) \leq 0$ для любого подмножества $A \subset \Omega^-$.

Очевидно (?), что $\Omega^+ = B$, $\Omega^- = B^c$.

Разложение Жордана. Если ν — ограниченный заряд на σ -алгебре подмножеств Ω , то

$$\nu = \nu^+ - \nu^-,$$

где ν^+ и ν^- — меры.

Обозначим: $\nu^+(A) = \nu(\Omega^+ \cap A)$, $\nu^-(A) = -\nu(\Omega^- \cap A)$.

Теперь перейдем к доказательству теоремы Радона–Никодима. В силу разложения Жордана достаточно доказать эту теорему для мер. Итак, ν — мера, непрерывная относительно меры μ . Рассмотрим все возможные неопределенные интегралы

$$A \rightarrow \int_A f d\mu$$

от неотрицательных измеримых функций f на Ω , для которых

$$\int_A f d\mu \leq \nu(A). \quad (*)$$

Очевидно (?), если $f = f_1$ и $f = f_2$ удовлетворяют (*), то и $f = \max\{f_1, f_2\}$, также удовлетворяет условию (*). Обозначим через F множество всех f , удовлетворяющих (*). Рассмотрим такую последовательность функций (f_n) , удовлетворяющих (*), что

$$\int_{\Omega} f_n d\mu \rightarrow \sup \left\{ \int_{\Omega} f d\mu : f \in F \right\}.$$

Супремум этой последовательности по теореме Лебега и по сделанному замечанию также удовлетворяет (*). Обозначим этот супремум f . Имеем:

$$\int_{\Omega} f d\mu = \sup \left\{ \int_{\Omega} f d\mu : f \in F \right\}.$$

Докажем, что

$$\int_A f d\mu = \nu(A) \quad \forall A \in \mathfrak{A}.$$

Предположим противное, и покажем, что тогда f можно увеличить с сохранением свойства (*), это приведет нас к противоречию с выбором f . Для этого достаточно рассмотреть новую меру

$$\lambda(A) = \nu(A) - \int_A f d\mu \quad \forall A \in \mathfrak{A},$$

предположить, что она не является нулевой, и доказать, что существует такая неотрицательная не равная всюду нулю функция g , что

$$\lambda(A) \geq \int_A g d\mu \quad \forall A \in \mathfrak{A}.$$

Доказав это, мы можем увеличить f , добавляя к нему g .

Предположим, что

$$\lambda(\Omega) = \varepsilon > 0. \tag{**}$$

Определим заряд

$$\rho(A) = \lambda(A) - \int_A \frac{\varepsilon}{2\mu(\Omega)} d\mu \quad \forall A \in \mathfrak{A}.$$

Согласно (**) заряд ρ на некоторых A принимает строго положительные значения, поэтому $\delta = \sup \rho > 0$. По разложению Хана существует такое измеримое множество Ω^+ , что $\rho(A) \geq 0$ для всех $A \subset \Omega^+$. Это означает, что

$$\lambda(A) \geq \int_A \frac{\varepsilon}{2\mu(\Omega)} d\mu \quad \forall A \subset \Omega^+.$$

Итак, мы можем положить

$$g = \frac{\varepsilon}{2\mu(\Omega)} \cdot I_{\Omega^+}.$$

Теорема Радона–Никодима по-видимому доказана. Заметим, что производная Радона–Никодима $\frac{d\nu}{d\mu}$ задана лишь с точностью до множества μ -нулевой меры, ее определение не использует порядок на числовой прямой (как обычная производная).

Упражнение. Докажите единственность $\frac{d\nu}{d\mu}$ с точностью до множества μ -нулевой меры.

Теперь мы используем ее для случая $\mu = \mathbb{P}$ и заряда $\nu(A) = \int_A \eta d\mu$. Оба заряда предполагаются заданными не на всей σ -алгебре событий \mathfrak{A} , а на ее σ -подалгебре \mathfrak{B} (в задаче регрессии это – наименьшая σ -алгебра, относительно которой измерима случайная величина (а в более общей ситуации случайный вектор) ξ). Итак, эта производная обозначается $\mathbb{E}(\eta|\mathfrak{B})$ или $\mathbb{E}_{\mathfrak{B}}\eta$. В частном случае, когда $\eta = I_A$, условное математическое ожидание называется условной вероятностью A при условии \mathfrak{B} и используется обозначение $\mathbb{P}(A|\mathfrak{B})$. В отличие от обычной вероятности (вероятность для нас уже привычное понятие!) условная вероятность – это неотрицательная случайная величина.

Упражнения.

- a. Докажите, что $0 \leq \mathbb{P}(A|\mathfrak{B}) \leq 1$ почти наверное.
- b. Докажите, что $\mathbb{P}\left(\sum_n A_n|\mathfrak{B}\right) = \sum_n \mathbb{P}(A_n|\mathfrak{B})$ почти наверное.
- c. Докажите, что если $A \in \mathfrak{B}$, то $\mathbb{P}(A|\mathfrak{B}) = \mathbb{P}(A)$ почти наверное.

Как определяется \mathfrak{B} в рассмотренных выше задачах регрессии? В дискретном случае $\xi = \sum_n a_n I_{\{\xi=a_n\}}$ в роли \mathfrak{B} выступает σ -подалгебра, состоящая из всевозможных объединений множеств вида $\{\xi = a_n\}$. В этом случае явно строится случайная величина $\mathbb{P}(A|\mathfrak{B})$, которая на каждом множестве $\{\xi = a_n\}$ принимает значение $\mathbb{P}(A|\xi = a_n)$. В непрерывном случае мы считаем, что $\Omega = \mathbf{R}^2$, на нем задана вероятность \mathbb{P} соотношением

$$\mathbb{P}(B) = \int_B p_{\xi,\eta}(x,y) dx dy.$$

В роли σ -алгебры \mathfrak{A} используется σ -алгебра всех борелевских множеств в \mathbf{R}^2 , а в роли \mathfrak{B} — σ -алгебра всех множеств вида $B \times \mathbf{R}$, где B пробегает все борелевские множества в \mathbf{R} .

Упражнение. Покажите, что функция

$$p_\eta(y|\xi=x) = \frac{p_{\xi,\eta}(x,y)}{\int_{-\infty}^{\infty} p_{\xi,\eta}(x,y) dy} \quad (*)$$

с точностью до значений на множестве меры нуль является функцией плотности вероятности $\mathbb{P}(\cdot|\mathfrak{B})$.

Неполное решение. По определению $\mathbb{P}(\cdot|\mathfrak{B})$ определена с точностью до множества меры нуль относительно первоначальной вероятности \mathbb{P} на \mathbf{R}^2 , задаваемой (?) функцией плотности $p_{\xi,\eta}$. Имея дело с очень абстрактно задаваемыми условным средним и условной вероятностью, трудно надеяться использовать какие-то вычисления. Надо идти от определения и показать, что функция (*) удовлетворяет определению условной вероятности. Вспомним, что

$$\mathbb{P}(A|\mathfrak{B}) := \mathbb{E}(I_A|\mathfrak{B}),$$

то есть любая функция, измеримая относительно σ -алгебры \mathfrak{B} , удовлетворяющая тождественному равенству

$$\int_B \mathbb{P}(A|\mathfrak{B}) d\mathbb{P} = \int_{AB} d\mathbb{P}$$

для всех $B \in \mathfrak{B}$.

Формулы умножения вероятностей, полной вероятности и Байеса для условной плотности. Мне неизвестен ответ на следующий

Вопрос. Пусть \mathfrak{B} — наименьшая σ -подалгебра, относительно которой измерима случайная величина ξ . Существует ли такая борелевская функция f , что

$$\mathbb{E}(\eta|\mathfrak{B}) = f(\xi) \text{ почти наверное?}$$

Важное значение имеет вопрос о существовании регулярной условной вероятности, то есть таких случайных величин $\pi(A)(\omega)$, что для каждого $\omega \in \Omega$ функция $A \rightarrow \pi(A)(\omega)$ является вероятностью, и

$$\mathbb{P}(A|\mathfrak{B}) = \pi(A)(\omega) \text{ почти наверное}$$

для любого A . Известны примеры, когда это неверно.

Условная плотность. Как ее интерпретировать с точки зрения условной вероятности? Распределение как интеграл по плотности. Иногда — это интеграл Римана (если плотность интегрируема по Риману, и множество также является хорошим).

Если имеется несколько вероятностей, то было бы удобно мыслить их как плотности относительно одной меры. Например, на каких событиях вероятность \mathbb{P}_1 ,больше вероятности \mathbb{P}_2 . Казалось бы, это верно, если $\mathbb{P}_1(A) \geq \mathbb{P}_2(A)$. Но на некоторых подмножествах A может быть справедливо обратное неравенство. В терминах плотности все просто. Что нужно, чтобы $\mathbb{P}_1(A)$ и $\mathbb{P}_2(A)$ были абсолютно непрерывны относительно одной меры. Всегда ли такая мера существует?

Упражнение 16. Покажите, что константой c , минимизирующей $\mathbb{E}|\xi - c|$ является не среднее значение ξ , а медиана ξ .

Часть 3. ПОСЛЕДОВАТЕЛЬНОСТИ СЛУЧАЙНЫХ ВЕЛИЧИН

Сходимость функций используется для замены сложных функций более простыми. Например, для вычисления числа e^x мы можем использовать разложение в ряд $e^x = 1+x+x^2/2!+\dots$. В данном случае мы приближаем экспоненциальную функцию полиномом. Очень удобно в этом случае, что мы всегда можем оценить величину ошибки, используя, например, остаточный член в разложении в форме Лагранжа или Коши.

Для случайных величин нам важно не столько приблизить случайные величины, сколько определить наилучшее приближение для данного ω . Но тогда мы можем использовать выбор: либо добиться успеха и сделать маленькой ошибку либо сделать маленькой вероятность ошибки. Тогда мы сможем надеяться, что нам повезло и ошибка в результате замены ξ на ξ_n для нашего ω невелика. Сходимость почти всюду означает, что мы можем быть уверены, что в нашем случае (для нашего ω мы в пределе получим истинное значенике). Сходимость по вероятности позволяет судить о том, что имеет практическое значение — о вероятности данной ошибки. Впрочем, сходимость почти наверное влечет сходимость по вероятности. Сходимость в среднем квадратическом тоже влечет сходимость по вероятности. Метод Монте-Карло позволяет использовать вероятностные идеи для приближенного вычисления кратных интегралов.

§9. Сходимость случайных величин

Формально случайные величины являются функциями, заданными на пространстве Ω случайных исходов со значениями в \mathbf{R} . Таким образом, для введения понятия сходимости последовательностей случайных величин мы можем воспользоваться аналогичными понятиями, известными нам для функций. Для последовательностей функций важнейшим является понятие равномерной сходимости, используя его, мы можем добиться приближения неизвестной функции f вычисляемой функцией f_n , добиваясь того, что величина ошибки равномерна мала, то есть меньше определенного числа, с которым мы согласны мириться, для всех значений аргумента. Дело в том, что функция f как правило является решением некоторой практической задачи и зависимость f от значений аргумента (например, времени) говорит о поведении системы.

Ситуация с последовательностями случайных величин несколько иная. Мы имеем дело со сходимостью $\xi_n(\omega)$ к $\xi(\omega)$ на разных $\omega \in \Omega$, но зависимость ξ от ω интереса не представляет, для нас важно, чтобы хорошее приближение получилось для того ω , который соответствует нашему случаю. Но так как найти или приблизить это ω не представляется возможным, мы готовы смириться с небольшой ошибкой не для всех ω , а лишь для ω с вероятностью очень близкой к 1. При этом мы надеемся на то, что невезение не окажется очень сильным и мы не попадем в ω , в котором приближение очень плохое.

Такой концепции соответствует *сходимость по вероятности* (в анализе она называется *сходимостью по мере*).

1. Определение. Последовательность ξ_n сходится к ξ по вероятности (пишем $\xi_n \xrightarrow{\mathbb{P}} \xi$), если для любого $\varepsilon > 0$

$$\mathbb{P}\{|\xi_n - \xi| > \varepsilon\} \rightarrow 0.$$

Свойства сходимости по вероятности

Если $\xi_n \xrightarrow{\mathbb{P}} \xi$, $\eta_n \xrightarrow{\mathbb{P}} \eta$, то

1) $\xi_n + \eta_n \xrightarrow{\mathbb{P}} \xi + \eta$,

2) $\xi_n \eta_n \xrightarrow{\mathbb{P}} \xi \eta$,

3) если дополнительно $\mathbb{P}\{\eta \neq 0\} = 1$, то $\xi_n / \eta_n \xrightarrow{\mathbb{P}} \xi / \eta$.

Доказательство этих свойств довольно интересно, но я приведу лишь доказательство 1):

$$\mathbb{P}\{|\xi_n + \eta_n - \xi - \eta| > \varepsilon\} \leq \mathbb{P}\{|\xi_n - \xi| > \varepsilon/2\} + \mathbb{P}\{|\eta_n - \eta| > \varepsilon/2\},$$

так как

$$\{|\xi_n + \eta_n - \xi - \eta| > \varepsilon\} \subset \{|\xi_n - \xi| > \varepsilon/2\} \cup \{|\eta_n - \eta| > \varepsilon/2\}.$$

Упражнение. Пусть $\mathbb{P}\{\xi_n = x_n\} = 1$, $\mathbb{P}\{\xi = x\} = 1$. Проверьте, что

$$x_n \rightarrow x \Leftrightarrow \xi_n \xrightarrow{\mathbb{P}} \xi.$$

Все эти свойства удобно также доказывать от противного, сводя рассмотрение к сходимости почти наверное, для которой все эти свойства немедленно следуют из свойств пределов последовательностей.

2. Определение. Мы говорим, что последовательность случайных величин (ξ_n) сходится к ξ *почти наверное* (пишем $\xi_n \xrightarrow{\text{П.Н.}} \xi$), если $\mathbb{P}\{\omega : \xi_n(\omega) \rightarrow \xi(\omega)\} = 1$. Из анализа известно, что из любой последовательности, сходящейся по вероятности, можно выделить подпоследовательность, сходящуюся почти наверное. Кроме того, ввиду конечности вероятностной меры, любая последовательность, сходящаяся почти наверное, сходится и по вероятности.

Мы предполагаем противное:

$$\xi_n \xrightarrow{\mathbb{P}} \xi, \quad \eta_n \xrightarrow{\mathbb{P}} \eta,$$

но $(\xi_n + \eta_n)$ не сходится по вероятности к $\xi + \eta$. Из определения следует, что для некоторого $\delta > 0$

$$\mathbb{P}\{|\xi_n + \eta_n - \xi - \eta| > \delta\} \not\rightarrow 0.$$

Тогда существует такое $\varepsilon > 0$, что для некоторой подпоследовательности n_k

$$\mathbb{P}\{|\xi_{n_k} + \eta_{n_k} - \xi - \eta| > \delta\} > \varepsilon \text{ для всех } k. \quad (*)$$

Выделим из n_k подпоследовательностей n_{k_m} так, что

$$\xi_{n_{k_m}} \xrightarrow{\text{П.Н.}} \xi, \quad \eta_{n_{k_m}} \xrightarrow{\text{П.Н.}} \eta.$$

Имеем:

$$\mathbb{P}\{\omega : \xi_{n_{k_m}}(\omega) \rightarrow \xi(\omega), \eta_{n_{k_m}}(\omega) \rightarrow \eta(\omega)\} = 1.$$

Следовательно,

$$\xi_{n_{k_m}} + \eta_{n_{k_m}} \xrightarrow{\text{П.Н.}} \xi + \eta,$$

но это по отмеченной выше связи двух сходимостей противоречит (*).

Аналогично можно доказать другие свойства сходимости по вероятности, но доказательства с помощью вероятностных неравенств более поучительны.

3. Теорема (Слабый закон больших чисел). *Пусть ξ_n – последовательность одинаково распределенных независимых случайных величин со средним и дисперсией (поэтому все ξ_n имеют одно и то же среднее, обозначим его m и одну и ту же дисперсию, обозначим ее σ^2). Тогда*

$$\frac{\xi_1 + \dots + \xi_n}{n} \xrightarrow{\mathbb{P}} m. \quad (1)$$

Доказательство несложно и основано на классическом неравенстве Чебышева. Воспользовавшись свойствами дисперсии, нетрудно подсчитать дисперсию левой части (1):

$$\mathbb{D}\left(\frac{\xi_1 + \dots + \xi_n}{n}\right) = \frac{\mathbb{D}(\xi_1 + \dots + \xi_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \rightarrow 0.$$

Точно так же

$$\mathbb{E}\left(\frac{\xi_1 + \dots + \xi_n}{n}\right) = \frac{\mathbb{E}(\xi_1 + \dots + \xi_n)}{n} = \frac{nm}{n} = m.$$

Поэтому нам достаточно доказать следующее неравенство Чебышева:

$$\mathbb{P}\{|\xi - \mathbb{E}\xi| > \varepsilon\} \leq \frac{\mathbb{D}\xi}{\varepsilon^2}. \quad (2)$$

Сначала доказывается неравенство для неотрицательной случайной величины η :

$$\mathbb{P}\{\eta > \varepsilon\} \leq \frac{\mathbb{E}\eta}{\varepsilon}.$$

После переноса ε в левую часть оно становится очевидным (?):

$$\varepsilon \mathbb{P}\{\eta > \varepsilon\} \leq \mathbb{E}\eta,$$

так как левая часть равна $\mathbb{E}\varepsilon I_{\{\eta > \varepsilon\}}$, среднее большей величины больше, а $\varepsilon I_{\{\eta > \varepsilon\}} < \eta$ (приверите при обоих значениях индикатора). Далее остается подставить $\eta = (\xi - \mathbb{E}\xi)$ и проверить, что $\{|\xi - \mathbb{E}\xi| > \varepsilon\} = \{(\xi - \mathbb{E}\xi)^2 > \varepsilon^2\}$. Подставив вместо ξ в (2) $\frac{\xi_1 + \dots + \xi_n}{n}$, мы получим

$$\mathbb{P}\left\{\left|\frac{\xi_1 + \dots + \xi_n}{n} - m\right| > \varepsilon\right\} \leq \frac{\sigma^2}{n}. \quad (3).$$

Это неравенство будет нам полезно и в других ситуациях.

4. Применение закона больших чисел — частотное определение вероятности

Мы начали наш курс с частотной интерпретации вероятности:

$$\lim_n \frac{n_A}{n} = \mathbb{P}(A),$$

где n_A — число экспериментов, в которых выпало событие A , n — общее число экспериментов, все эксперименты проводятся в одинаковых условиях и каждый раз наблюдается одно и то же событие A . Как построить вероятностную модель этого процесса. Событие A , наблюданное в i -м эксперименте, мы обозначим A_i . Проведение эксперимента в одинаковых условиях можно интерпретировать как независимость в совокупности событий A_i (результат одного эксперимента не влияет на исход других), так как наблюдается одно и то же событие, то все вероятности $\mathbb{P}(A_i)$ должны быть равны одному и тому же числу — вероятности $\mathbb{P}(A)$ события A .

Итак, случайные величины I_{A_i} имеют одно и то же распределение и независимы, $\mathbb{E}(A_i) = \mathbb{P}(A)$. Согласно закону больших чисел

$$\frac{n_A}{n} = \frac{I_{A_1} + I_{A_2} + \dots + I_{A_n}}{n} \xrightarrow{\mathbb{P}} \mathbb{P}(A).$$

Конечно, из этого рассуждения лишь следует, что с вероятностью близкой к 1 число $\frac{n_A}{n}$ близко к $\mathbb{P}(A)$, но не следует, что именно у нас $\frac{n_A}{n}$ обязательно сходится к $\mathbb{P}(A)$. Этот факт следует из сходимости не по вероятности, а почти наверное, т. е. для всех ω , за исключением ω из множества с вероятностью 0. Закон больших чисел со сходимостью почти наверное доказан Колмогоровым (при наших предположениях и даже более общих), но мы это доказательство проходить пока не будем.

Упражнение 17. Как известно, случайная величина с распределением Стьюдента имеет вид

$$t_n = \frac{\xi}{\sqrt{\frac{1}{n}(\xi_1^2 + \dots + \xi_n^2)}}.$$

Докажите, что если числитель в этой формуле один и тот же для всех n , то

$$t_n \xrightarrow{\mathbb{P}} \xi.$$

§10. Применение закона больших чисел — метод Монте-Карло

Метод Монте-Карло играет большую и все возрастающую роль в многих компьютерных вычислениях. В нем случайность моделируется искусственно. Используется имеющийся в многих программных пакетах датчик случайных чисел. Предполагается, что этот датчик выплевывает числа x_1, x_2, \dots, x_n , которые являются наблюдениями независимых случайных величин с равномерным распределением на отрезке $[0, 1]$. Это должно означать, что для любого boreлевского множества $B \subset [0, 1]$ доля чисел x_i , попавших в B , должна стремится к длине B , доля всех пар (x_i, x_{i+1}) , попавших в подмножество $B \subset [0, 1] \times [0, 1]$, сходится к площади B , аналогичные утверждения должны быть верны для последовательных троек и объемов, для последовательных четверок и четырехмерных объемов, и т. д. На практике идеальную последовательность с таким свойствами получить невозможно, и строится последовательность не случайных чисел, а чисел, называемых 'псевдослучайными'. Алгоритмы для создания псевдослучайных чисел создавались специалистами по теории чисел. Более того, они проверялись на большом числе практических вычислений, некоторые из них были в конце концов отвергнуты (см. замечательную брошюру И.М. Соболь...) Тем не менее, в теоретических рассуждениях мы предполагаем, что независимость и одинаковая равномерная распределенность на $[0, 1]$ имеют место. Таким образом, мы считаем, что числа x_i — наблюдения случайных величин ξ_i , которые являются независимыми и с равномерным распределением. Мы обозначим также через ξ одну эталонную случайную величину с равномерным распределением на $[0, 1]$. Рассмотрим теперь функцию f с хорошими свойствами на $[0, 1]$, тогда согласно закону больших чисел

$$\frac{f(\xi_1) + f(\xi_2) + \dots + f(\xi_n)}{n} \xrightarrow{\mathbb{P}} \mathbb{E}f(\xi) = \int_0^1 f(x)dx.$$

Последнее равенство следует из формулы замены переменной

$$\mathbb{E}f(\xi) = \int f(x)p_\xi(x)dx,$$

причем в случае равномерного распределения на $[0, 1]$

$$p_\xi(x) = I_{[0,1]}(x).$$

Так как x_i мы считаем наблюденными значениями случайных величин ξ_i , то мы считаем, что с вероятностью, близкой к 1, число

$$\frac{f(x_1) + f(x_2) + \dots + f(x_n)}{n}$$

близко к

$$\int_0^1 f(x)dx.$$

Можно даже оценить меру близости вероятности (к 1) и числа. Давайте потребуем, чтобы вероятность нужной нам близости была бы не меньше чем 0.999. Для оценки мы будем использовать простейшее неравенство Чебышева. Но чтобы его использовать, мы должны оценить $\sigma^2 = \mathbb{D}f(\xi)$. Пусть $|f| \leq C$. Вы можете хотя бы на примере двухзначных случайных величин убедиться, что $\mathbb{D}f(\xi) \leq C^2$. Имеем:

$$\mathbb{P} \left\{ \left| \frac{f(\xi_1) + \dots + f(\xi_n)}{n} - \int_0^1 f(x)dx \right| > \varepsilon \right\} \leq \frac{C^2}{n\varepsilon^2}$$

Если мы будем считать допустимой ошибкой порядка одной тысячной от C , то правая часть будет меньше 0,001 при $n = 10^9$. Замечу, что имеются экспоненциальные оценки (например, неравенство Бернштейна), которые позволяют существенно уменьшить правую часть. Итак, чтобы получить приличную точность с приличной вероятностью, нужен очень большой объем вычислений. В известных вам приближенных методах вычисления интеграла, накладываются условия на f не только ограниченности, но и непрерывности и дифференцируемости. Здесь этого не нужно. Но ясно, что для одномерного интеграла все это не интересно — для получения той же точности (при более ограничительных условиях) нужно вычислять функцию f не более чем в тысяче точек. Так когда-то до создания компьютеров и считали. Однако метод Монте-Карло можно использовать и при вычислении многомерных интегралов от функции $f(x_1, x_2, \dots, x_r)$ по r -мерному кубу $[0, 1] \times [0, 1] \times \dots \times [0, 1]$. Число точек в классических методах вычисления интегралов растет как n^r , стомерный интеграл такими методами не вычислишь. А для вычисления методом Монте-Карло для получения той же точности нужно то же число точек, как и в одномерном случае. Но каждая точка в области интегрирования будет теперь задаваться не одним псевдослучайным числом, а r -кой псевдослучайных чисел. Еще эффективнее метод Монте-Карло выглядит для более сложных областей интегрирования чем r -мерный куб. Обычно такая область задается несколькими неравенствами (например, $x_1^2 + \dots + x_r^2 \leq 1$). Заменой переменных мы загоняем область интегрирования в единичный r -мерный куб, но в интегральной сумме

$$\frac{f(x_1^{(1)}, x_2^{(1)}, \dots, x_r^{(1)}) + \dots + f(x_1^{(n)}, x_2^{(n)}, \dots, x_r^{(n)})}{n}$$

мы берем лишь те слагаемые, которые заданы на точках $(x_1^{(i)}, \dots, x_r^{(i)})$ из области интегрирования.

Метод Монте-Карло используется не только для вычисления интегралов, но и для численного решения дифференциальных уравнений, для моделирования молекул с использованием методов квантовой химии. Он требует гигантского объема вычислений, но в многих ситуациях является единственным возможным.

§11. Усиленный закон больших чисел

Теорема (Усиленный закон больших чисел Колмогорова) *Пусть ξ_i — последовательность независимых одинаково распределенных случайных величин, имеющих среднее значение. Обозначим $\mathbb{E}(\xi_i) = m$. Тогда*

$$\frac{\xi_1 + \xi_2 + \dots + \xi_n}{n} \xrightarrow{\text{п.н.}} m. \quad (1)$$

Замечу, что Колмогоров доказал и обратное утверждение: если для независимых одинаково распределенных случайных величин ξ_n имеет место

$$\frac{\xi_1 + \xi_2 + \dots + \xi_n}{n} \xrightarrow{\text{п.н.}} a,$$

где a — константа, то случайные величины ξ_n имеют среднее, и a совпадает со средним. Этот факт мы доказывать не будем.

Замечание. Отмечу, что сходимость почти наверное (или почти всюду) отнюдь не означает сходимость всюду. Может быть расходимость на некотором событии, но вероятность этого события равна нулю. Значит, мы должны научиться как-то выделять события нулевой вероятности и доказывать равенство нулю вероятности для таких событий.

Приведу любопытный *пример*. Рассмотрим последовательность независимых случайных величин ε_n , принимающих с вероятностью $1/2$ значения -1 и 1 . Ясно, что такая последовательность удовлетворяет условиям теоремы. Но чтобы все было убедительно, я должен представить вам вероятностное пространство и на нем соответствующую последовательность случайных величин. Оказывается, это легко сделать. Возьмем в качестве вероятностного пространства

интервал $[0, 1)$ с мерой Лебега, в частности, вероятность интервала — это длина этого интервала. И рассмотрим последовательность случайных величин, которые в данном случае — измеримые функции:

$$\begin{aligned}\varepsilon_1 &= \begin{cases} 1, & \text{если } x \in [0, 1/2), \\ -1, & \text{если } x \in [1/2, 1), \end{cases} \quad \varepsilon_2 = \begin{cases} 1, & \text{если } x \in [0, 1/4) \cup [1/2, 3/4), \\ -1, & \text{если } x \in [1/4, 1/2) \cup [3/4, 1), \end{cases} \\ \varepsilon_3 &= \begin{cases} 1, & \text{если } x \in [0, 1/8) \cup [1/4, 3/8) \cup [1/2, 5/8) \cup [3/4, 7/8), \\ -1, & \text{если } x \in [1/8, 1/4) \cup [3/8, 1/2) \cup [5/8, 3/4) \cup [7/8, 1), \end{cases} \quad \dots\end{aligned}$$

Думаю, что вы сообразили сами, как строятся случайные величины $\varepsilon_4, \varepsilon_5$ и т. д. Нетрудно проверить, и вам придется это сделать самим, что случайные величины ε_n независимы (пропритесть независимость для пар $\varepsilon_n, \varepsilon_m$, где $m \neq n$, потом для троек, думаю, этого хватит). По теореме Колмогорова, которую мы с вами докажем,

$$\frac{\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_n}{n} \xrightarrow{\text{п.н.}} 0.$$

Но это верно не для всех случайных исходов $\omega \in [0, 1]$. например, при $\omega = 0, \varepsilon_n(1) = 1$ для всех n и предел будет равен 1, а не 0. Легко видеть, что лишь конечное число нулей будет у нашей последовательности в любой точке $\omega = k/2^n$, поэтому предел также равен 1. А вот придумать такое (нерациональное) ω , что предел равен 0, я не могу. Тем более, не могу вычислить меру Лебега таких точек и доказать, что она равна 1. Так что, убедиться в справедливости теоремы 'на примерах' не получается. Итак, будьте внимательны, а вдруг теорема великого Колмогорова неверна.

В замечательной книге Пойя 'Математика и правдоподобные рассуждения' в одной из глав используется эпиграф: 'Убедившись, что теорема верна, начинаем ее доказывать. Традиционный профессор математики.'

К нашей ситуации этот эпиграф неприменим.

Кстати, последовательность построенных функций, если добавить к ней функцию, тождественно равную 1, называется системой Радемахера, она ортонормирована, но не полна. Но если взять все конечные произведения таких функций Радемахера, то получим ортонормированный базис Уолша в пространстве $L^2[0, 1)$. Этот базис используется в разных прикладных задачах, в частности, в задачах сжатия информации.

План доказательства. Сначала мы докажем усиленный закон больших чисел в иной ситуации. Если бы у одинаково распределенных случайных величин ξ_n существовали дисперсии $\mathbb{D}(\xi_n)$, которые также должны быть равны, то

$$\sum_n \frac{\mathbb{D}(\xi_n)}{n^2} < \infty. \tag{2}$$

Мы сделаем следующее: заменим в условиях теоремы независимые одинаково распределенные случайные величины ξ_n , у которых существует среднее m ; на независимые случайные величины η_n , у которых существуют дисперсии и сходится ряд

$$\sum_n \frac{\mathbb{D}(\eta_n)}{n^2} < \infty. \tag{2'}$$

а) Мы докажем, что для последовательности независимых случайных величин (η_n) , удовлетворяющей условию (2'), ряд

$$\sum_n \frac{\eta_n - \mathbb{E}\eta_n}{n}. \tag{3}$$

сходится почти наверное.

b) Отсюда выводится

$$\frac{(\eta_1 - \mathbb{E}\eta_1) + (\eta_2 - \mathbb{E}\eta_2) + \dots + (\eta_n - \mathbb{E}\eta_n)}{n} \xrightarrow{\text{п.н.}} 0. \quad (4)$$

c) Далее мы возвращаемся к формулировке теоремы и показываем, что после незначительного изменения случайные величины ξ_n из условий теоремы будут удовлетворять условию (2). Из этого факта мы выведем доказательство теоремы.

Для доказательства а) нам нужно провести большую дополнительную работу.

i) Итак, мы хотим доказать, что при выполнении условия (2')

$$\mathbb{P} \left\{ \text{ряд } \sum_n \frac{\eta_n - \mathbb{E}(\eta_n)}{n} \text{ расходится} \right\} = 0. \quad (3')$$

Однако сходимость для почти всех точек не означает сходимость для всех точек (постройте пример, когда нет сходимости на множестве всех рациональных точек).

Поэтому нужно научиться выделять события вероятности нуль, и доказывать, что это — событие вероятности нуль.

ii) Кроме того, нужен какой-то критерий сходимости п.н. ряда $\sum_n \eta_n$ случайных величин, т. е. сходимости числового ряда $\sum_n \eta_n(\omega)$ за исключением ω из множества меры нуль.

iii) Для этого нам понадобится, как это не странно, новый для вас критерий сходимости числового ряда.

Перейдем к пункту i) доказательства а).

i) **Лемма Бореля – Кантелли.** Пусть (A_n) — последовательность событий, $\sum_n \mathbb{P}(A_n) < \infty$, тогда

$$\mathbb{P}\{\omega : \omega \in A_n \text{ бесконечное число раз}\} = 0. \quad (5)$$

Доказательство. Проверим, что

$$A = \{\omega : \omega \in A_n \text{ бесконечное число раз}\} = \bigcap_n \bigcup_{m \geq n} A_m.$$

Действительно, если $\omega \in A_n$ бесконечное число раз, то для любого n найдется такое $m \geq n$, что $\omega \in A_m$, но тогда $\omega \in \bigcup_{m \geq n} A_m$ для любого n . Так как остаток сходящегося ряда $\sum_n \mathbb{P}(A_n)$ сходится к 0, мы получаем

$$\mathbb{P}(A) \leq \mathbb{P} \left\{ \bigcup_{m \geq n} A_m \right\} \leq \sum_{m \geq n} \mathbb{P}(A_m) \rightarrow 0.$$

Упражнение. Докажите, что если события A_n независимы в совокупности, то из (4) следует $\sum_n \mathbb{P}(A_n) < \infty$.

ii) **Лемма 1. Критерий типа Коши п.н. сходимости ряда из случайных величин.** Ряд из случайных величин $\sum_n \eta_n$ сходится п. н. тогда и только тогда, когда для любых $\varepsilon > 0$ и $\delta > 0$ существует $N \in \mathbf{N}$ такое что для любого натурального $r > 0$

$$\mathbb{P} \left\{ \omega : \max_{s \leq r} \left| \sum_{i=N+1}^{N+s} \eta_i(\omega) \right| > \delta \right\} \leq \varepsilon. \quad (**)$$

Теперь приведем нужный нам критерий сходимости числового ряда, который представляет собой нечто среднее между критерием Коши сходимости ряда и свойством сходимости ряда, сходящегося абсолютно.

iii) Ряд $\sum_n x_n$ сходится тогда и только тогда, когда существуют сходящийся ряд $\sum_k \delta_k$, где $\delta_k \geq 0$ для всех k , и возрастающая последовательность натуральных чисел $n_1 < n_2 < \dots < n_k < \dots$, такие, что для любого k

$$\max_{n_k < s \leq n_{k+1}} \left| \sum_{i=n_k+1}^s x_i \right| \leq \delta_k. \quad (*)$$

Для наших целей нам достаточно знать, что $(*)$ влечет сходимость ряда $\sum_n x_n$. Проверим выполнение условий критерия Коши: пусть дано произвольное $\varepsilon > 0$, выберем такое k , что

$$\sum_{s \geq k} \delta_s < \varepsilon.$$

Теперь в критерии Коши мы выберем $N = n_k$. Проверим нужное равенство для куска ряда от $N + 1$ до $N + r$. Пусть $l \geq k$ таково, что $N + r > n_l$. Мы имеем нужную нам оценку:

$$\left| \sum_{i=N+1}^{N+r} x_i \right| \leq \left| \sum_{i=n_k+1}^{n_{k+1}} x_i \right| + \dots + \left| \sum_{i=n_{l-1}+1}^{n_l} x_i \right| + \left| \sum_{i=n_l+1}^{N+r} x_i \right| \leq \sum_{j=k}^l \delta_j \leq \varepsilon.$$

Перейдем к доказательству леммы. Так как $\varepsilon > 0$ и $\delta > 0$ могут быть выбраны сколь угодно малыми, мы выберем последовательности ε_k и δ_k , сходящиеся к нулю достаточно быстро, а по ним последовательность натуральных чисел $n_k = N$ из соотношения $(**)$. Мы должны убить двух зайцев: с одной стороны сделать сходящимся ряд $\sum_k \delta_k$, а с другой стороны сделать сходящимся ряд из вероятностей

$$\mathbb{P} \left\{ \omega : \max_{n_k \leq s \leq n_{k+1}} \left| \sum_{i=n_k+1}^s \eta_i(\omega) \right| > \delta_k \right\} \leq \varepsilon_k,$$

для этого мы выбираем сходящимся ряд $\sum_k \varepsilon_k$. Из условия леммы следует, что это возможно.

Обозначим

$$A_k = \left\{ \omega : \max_{n_k \leq s \leq n_{k+1}} \left| \sum_{i=n_k+1}^s \eta_i(\omega) \right| > \delta_k \right\}.$$

Согласно лемме Бореля — Кантелли, мера всех ω , принадлежащих бесконечному числу элементов последовательности (A_k) , равна нулю. С вероятностью 1 элементарный исход ω не попадает в A_k , начиная с некоторого $k = k(\omega)$. Но тогда начиная с $k(\omega)$ для ω верно обратное неравенство:

$$\max_{n_k \leq s \leq n_{k+1}} \left| \sum_{i=n_k+1}^s \eta_i(\omega) \right| \leq \delta_k,$$

где ряд из δ_k сходится. В результате числовой ряд $\sum_i \eta_i(\omega)$ удовлетворяет нашему критерию сходимости, а следовательно, сходится.

Необходимость в лемме мы использовать не будем, ее доказательство является упражнением.

Теперь докажем а).

Теорема о п.н. сходимости ряда независимых случайных величин. Если последовательность независимых случайных величин η_n удовлетворяет условию

$$\sum_n \mathbb{D}(\eta_n) < \infty,$$

то ряд

$$\sum_n (\eta_n - \mathbb{E}(\eta_i))$$

сходится почти наверное.

Доказательство состоит в проверке условий в лемме 1, для чего достаточно доказать следующее замечательное

Неравенство Колмогорова. Пусть случайные величины $\zeta_i, i \leq r$, независимы и имеют дисперсию. Тогда

$$\mathbb{P} \left\{ \max_{s \leq r} \left| \sum_{i=1}^s [\zeta_i - \mathbb{E}(\zeta_i)] \right| > \delta \right\} \leq \frac{\sum_{i=1}^r \mathbb{D}(\zeta_i)}{\delta^2}.$$

Перепишем неравенство в виде

$$\delta^2 \mathbb{P} \left\{ \max_{s \leq r} \left| \sum_{i=1}^s [\zeta_i - \mathbb{E}(\zeta_i)] \right| > \delta \right\} \leq \sum_{i=1}^r \mathbb{D}(\zeta_i). \quad (***)$$

Событие A в левой части неравенства мы представим в виде суммы несовместных событий

$$A_1 = \{|\zeta_1 - \mathbb{E}(\zeta_1)| > \delta\},$$

$$A_2 = \{|\zeta_1 - \mathbb{E}(\zeta_1)| \leq \delta, |[\zeta_1 - \mathbb{E}(\zeta_1)] + [\zeta_2 - \mathbb{E}(\zeta_2)]| > \delta\}, \dots,$$

$$A_r = \{\dots, |[\zeta_1 - \mathbb{E}(\zeta_1)] + \dots + [\zeta_{r-1} - \mathbb{E}(\zeta_{r-1})]| \leq \delta,$$

$$|[\zeta_1 - \mathbb{E}(\zeta_1)] + \dots + [\zeta_r - \mathbb{E}(\zeta_r)]| > \delta\}.$$

Используя свойство дисперсии (дисперсия суммы независимых случайных величин равна сумме дисперсий), но в обратную сторону, мы перепишем правую часть (***):

$$\begin{aligned} \sum_{i=1}^r \mathbb{D}(\zeta_i) &= \mathbb{D} \left(\sum_{i=1}^r \zeta_i \right) = \int_{\Omega} \left\{ \sum_{i=1}^r [\zeta_i - \mathbb{E}(\zeta_i)] \right\}^2 d\mathbb{P} \geq \\ &\geq \int_A = \sum_{s=1}^r \int_{A_s} = \sum_{s=1}^r \int_{A_s} \left\{ \sum_{i=1}^s [\zeta_i - \mathbb{E}(\zeta_i)] + \sum_{i=s+1}^r [\zeta_i - \mathbb{E}(\zeta_i)] \right\}^2 d\mathbb{P} = \\ &= \sum_{s=1}^r \int_{A_s} \left\{ \sum_{i=1}^s [\zeta_i - \mathbb{E}(\zeta_i)] \right\}^2 + \left\{ \sum_{i=s+1}^r [\zeta_i - \mathbb{E}(\zeta_i)] \right\}^2 + \\ &\quad + 2 \left\{ \sum_{i=1}^s [\zeta_i - \mathbb{E}(\zeta_i)] \right\} \left\{ \sum_{i=s+1}^r [\zeta_i - \mathbb{E}(\zeta_i)] \right\} d\mathbb{P}. \end{aligned}$$

Теперь рассмотрим отдельно слагаемые в интеграле \int_{A_s} . Согласно определению события A_s имеем:

$$\int_{A_s} \left\{ \sum_{i=1}^s [\zeta_i - \mathbb{E}(\zeta_i)] \right\}^2 d\mathbb{P} \geq \delta^2 \mathbb{P}(A_s).$$

Очевидно, что

$$\int_{A_s} \left\{ \sum_{i=s+1}^r [\zeta_i - \mathbb{E}(\zeta_i)] \right\}^2 d\mathbb{P} \geq 0.$$

Далее,

$$\begin{aligned} & \int_{A_s} \left\{ \sum_{i=1}^s [\zeta_i - \mathbb{E}(\zeta_i)] \right\} \left\{ \sum_{i=s+1}^r [\zeta_i - \mathbb{E}(\zeta_i)] \right\} d\mathbb{P} = \int_{\Omega} I_{A_s} \left\{ \sum_{i=1}^s [\zeta_i - \mathbb{E}(\zeta_i)] \right\} \left\{ \sum_{i=s+1}^r [\zeta_i - \mathbb{E}(\zeta_i)] \right\} d\mathbb{P} = \\ & = \mathbb{E} \left(I_{A_s} \left\{ \sum_{i=1}^s [\zeta_i - \mathbb{E}(\zeta_i)] \right\} \left\{ \sum_{i=s+1}^r [\zeta_i - \mathbb{E}(\zeta_i)] \right\} \right) = \mathbb{E} \left(I_{A_s} \left\{ \sum_{i=1}^s [\zeta_i - \mathbb{E}(\zeta_i)] \right\} \right) \mathbb{E} \left\{ \sum_{i=s+1}^r [\zeta_i - \mathbb{E}(\zeta_i)] \right\} = 0. \end{aligned}$$

Прокомментируем эту выкладку. По определению события A_s случайная величина

$$I_{A_s} \left\{ \sum_{i=1}^s [\zeta_i - \mathbb{E}(\zeta_i)] \right\}$$

является функцией случайных величин ζ_i , где $i \leq s$, в то же время второй множитель под знаком \mathbb{E} — функция случайных величин ζ_i , где $i \geq s+1$. (На всякий случай проверьте сами, что по известным числам ζ_i , где $i \leq s$, однозначно восстанавливается $I_{A_s} \left\{ \sum_{i=1}^s [\zeta_i - \mathbb{E}(\zeta_i)] \right\}$.) Таким образом, оба множителя независимы согласно теореме 4 §6 (как функции независимых наборов случайных величин), среднее значение произведения равно произведению средних, но среднее значение второго множителя равно 0.

Если просуммировать все полученные оценки, мы получим

$$\sum_{i=1}^r \mathbb{D}(\zeta_i) \geq \delta^2 \sum_{s=1}^r \mathbb{P}(A_s) = \delta^2 \mathbb{P}(A).$$

Теперь, если выполнены условия теоремы о п.н. сходимости ряда, мы легко можем проверить условия критерия Коши п.н. сходимости ряда. А именно, мы выбираем такое N , что $\sum_{n=N+1}^{\infty} \mathbb{D}(\eta_n) < \delta^2 \varepsilon$ и подставляем в неравенство Колмогорова $\zeta_1 = \eta_{N+1}, \dots, \zeta_r = \eta_{N+r}$ (проверьте!).

Следствие 1. Если последовательность независимых случайных величин η_i , удовлетворяет условию

$$\sum_n \frac{\mathbb{D}(\eta_n)}{n^2} < \infty,$$

то ряд

$$\sum_n \frac{\eta_n - \mathbb{E}\eta_n}{n}$$

сходится почти наверное.

Дальше нам понадобится сходимость по Чезаро: последовательность s_n называется сходящейся по Чезаро к s , если

$$\frac{(s_1 - s) + (s_2 - s) + \dots + (s_n - s)}{n} \rightarrow 0.$$

Оправданием этого понятия является следующая

Теорема Чезаро. Если $s_n \rightarrow s$ (в обычном смысле), то s_n сходится к s по Чезаро.

Доказательство очевидно. Начиная с $n > N$ верно $|s_n - s| \leq \varepsilon$, в то же время

$$\begin{aligned} & \frac{(s_1 - s) + (s_2 - s) + \dots + (s_n - s)}{n} = \\ & = \frac{(s_1 - s) + (s_2 - s) + \dots + (s_N - s)}{n} + \frac{(s_{N+1} - s) + \dots + (s_n - s)}{n}. \end{aligned}$$

В этом представлении первое слагаемое сходится к нулю, так как числитель фиксирован, а второе слагаемое меньше $\frac{n-N}{n}\varepsilon$.

Сейчас мы установим b), для этого введем обозначение $x_n = \eta_n(\omega) - \mathbb{E}\eta_n$, рассмотрим те элементарные исходы ω , для которых ряд $\sum_n x_n/n$ сходится к s и покажем, что тогда

$$\frac{x_1 + x_2 + \dots + x_n}{n} \rightarrow 0. \quad (5)$$

Обозначим

$$s_n = \frac{x_1}{1} + \frac{x_2}{2} + \dots + \frac{x_n}{n} \rightarrow s.$$

Тогда левая часть (5) имеет вид

$$\frac{s_1 + 2(s_2 - s_1) + \dots + n(s_n - s_{n-1})}{n} = \frac{ns_n - s_1 - \dots - s_{n-1}}{n}. \quad (6)$$

Так как $s_n \rightarrow s$, из теоремы о том, что обычная сходимость влечет сходимость по Чезаро, следует, что $\frac{s_1 + \dots + s_{n-1}}{n} \rightarrow s$. Отсюда следует, что выражение в (6) сходится к 0.

Итак, мы по существу доказали следующее

Следствие 2. Если последовательность независимых случайных величин η_i удовлетворяет условию

$$\sum_n \frac{\mathbb{D}(\eta_n)}{n^2} < \infty,$$

то

$$\frac{(\eta_1 - \mathbb{E}\eta_1) + (\eta_2 - \mathbb{E}\eta_2) + \dots + (\eta_n - \mathbb{E}\eta_n)}{n} \xrightarrow{\text{п.н.}} 0. \quad (9)$$

c) Теперь мы наконец перейдем к доказательству основной теоремы, т.е. (1). рассмотрим случай одинаково распределенных случайных величин ξ_n , у которых существует среднее. Напомним, что для любой измеримой функции f

$$\int_{\mathbf{R}} f(x) \mathbb{P}_{\xi_1}(x) = \dots = \int_{\mathbf{R}} f(x) \mathbb{P}_{\xi_n}(x) = \dots,$$

и что интегрируемость означает для интеграла Лебега абсолютную интегрируемость, т.е.

$$\mathbb{E}|\xi_1| = \int_{\mathbf{R}} |x| \mathbb{P}_{\xi_1}(x) < \infty.$$

Рассмотрим новые случайные величины

$$\eta_n = \xi_n I_{\{|\xi_n| \leq n\}}.$$

Как функции независимых случайных величин, новые случайные величины тоже независимы, но их распределения не обязательно одинаковы. Покажем, что они удовлетворяют условиям пункта a):

$$\sum_n \frac{\mathbb{E}(\eta_n)^2}{n^2} < \infty. \quad (7)$$

Мы выразим эту сумму через (совпадающие) распределения случайных величин ξ_i , т. е., через распределение ξ_1 .

$$\begin{aligned} \sum_n \frac{\mathbb{E}(\eta_n)^2}{n^2} &= \sum_n \int_{|x| \leq n} |x|^2 \frac{1}{n^2} d\mathbb{P}_{\xi_1}(x) = \\ &= \sum_n \sum_{1 \leq i \leq n} \int_{|x| \leq i} = \end{aligned}$$

$$\begin{aligned}
&= \sum_n \int_{n-1 < |x| \leq n} |x|^2 \left(\frac{1}{n^2} + \frac{1}{(n+1)^2} + \dots \right) d\mathbb{P}_{\xi_1} \leq \\
&\leq \sum_n \int_{n-1 < |x| \leq n} |n|^2 \frac{C}{n} d\mathbb{P}_{\xi_1} \leq \int_{\mathbf{R}} C(|x|+1) d\mathbb{P}_{\xi_1} \leq \text{Const}.
\end{aligned}$$

Так как дисперсии неотрицательны,

$$\mathbb{E}(\eta_n)^2 \geq [\mathbb{E}(\eta_n)]^2.$$

Поэтому из $\sum_n \frac{\mathbb{E}(\eta_n)^2}{n^2} < \infty$ следует $\sum_n \frac{\mathbb{D}(\eta_n)}{n^2} < \infty$.

Таким образом, условия следствия 2 выполнены, и для последовательности η_n справедлив усиленный закон больших чисел в формулировке следствия 2. Дальнейшая наша задача — доказать справедливость этой формулировки также для первоначальных случайных величин ξ_n , а отсюда вывести справедливость и формулировки (1).

Докажем, что

$$\sum_n \mathbb{P}\{\xi_n \neq \eta_n\} < \infty. \quad (10)$$

Здесь опять выгодно выразить сумму через интеграл по распределению ξ_1 .

$$\begin{aligned}
\sum_n \mathbb{P}\{\xi_n \neq \eta_n\} &= \sum_n \int_{(n, \infty)} d\mathbb{P}_{\xi_1} = \\
\sum_n \int_{\mathbf{R}} I_{(n, \infty)} d\mathbb{P}_{\xi_1} &= \int_{\mathbf{R}} \sum_n I_{(n, \infty)}(x) d\mathbb{P}_{\xi_1} \leq \\
&\leq \int_{\mathbf{R}} |x| d\mathbb{P}_{\xi_1} < \infty.
\end{aligned}$$

В итоге, опять по лемме Бореля-Кантелли, с вероятностью 1 случайные величины $\xi_n(\omega)$ совпадают с $\eta_n(\omega)$, начиная с некоторого случайного $n = n(\omega)$. Причем нетрудно видеть, что для всех таких $\omega \in \tilde{\Omega}$ (где $\mathbb{P}\tilde{\Omega} = 1$) из сходимости

$$\frac{(\eta_1(\omega) - \mathbb{E}\eta_1) + (\eta_2(\omega) - \mathbb{E}\eta_2) + \dots + (\eta_n(\omega) - \mathbb{E}\eta_n)}{n} \rightarrow 0.$$

следует сходимость

$$\frac{(\xi_1(\omega) - \mathbb{E}\eta_1) + (\xi_2(\omega) - \mathbb{E}\eta_2) + \dots + (\xi_n(\omega) - \mathbb{E}\eta_n)}{n} \rightarrow 0.$$

Действительно, числители в обеих дробях отличаются лишь на фиксированное конечное число слагаемых, а знаменатель сходится к нулю, поэтому разность дробей сходится к нулю. Итак, из

$$\frac{(\eta_1(\omega) - \mathbb{E}\eta_1) + (\eta_2 - \mathbb{E}\eta_2) + \dots + (\eta_n - \mathbb{E}\eta_n)}{n} \xrightarrow{\text{п.н.}} 0.$$

следует

$$\frac{(\xi_1 - \mathbb{E}\eta_1) + (\xi_2 - \mathbb{E}\eta_2) + \dots + (\xi_n - \mathbb{E}\eta_n)}{n} \xrightarrow{\text{п.н.}} 0. \quad (11)$$

Для окончательного доказательства теоремы нам осталось заменить в (11) $\mathbb{E}\eta_i$ на $\mathbb{E}\xi_i$ для всех i , а для этого достаточно доказать сходимость

$$\frac{(\mathbb{E}\xi_1 - \mathbb{E}\eta_1) + (\mathbb{E}\xi_2 - \mathbb{E}\eta_2) + \dots + (\mathbb{E}\xi_n - \mathbb{E}\eta_n)}{n} \longrightarrow 0. \quad (12)$$

Но вспомним, что $\mathbb{E}\xi_n = \mathbb{E}\xi_1$. Соотношение (12) следует из сходимости

$$\mathbb{E}\eta_n = \int_{\Omega} \xi_1 I_{\{|\xi_1| \leq n\}} d\mathbb{P} \rightarrow \int_{\Omega} \xi_1 d\mathbb{P} = \mathbb{E}\xi_1,$$

которое само следует из теоремы Лебега для всюду сходящейся последовательности $\xi_1 I_{\{|\xi_1| \leq n\}}$ (ограниченной интегрируемой функцией $|\xi_1|$), и из теоремы Чезаро. Поэтому, чтобы прийти к формуле (1), нам достаточно перенести t в правую часть.

Теорема Леви и ее следствия для интегралов по распределению Распределение – ограниченная мера, поэтому имеется сходимость интегралов по распределению от ограниченных и даже интегрируемых функций. Дифференцирование под знаком интеграла.

Как записываются интегралы по распределению — $\int dF_\xi$, $d\int \mathbb{P}_\xi$.

§12. Слабая сходимость распределений

В теории вероятностей (в отличие от изучаемой в курсе функционального анализа теории меры) кроме сходимости функций (они в теории вероятностей называются случайными величинами, а точки, на которых они заданы, называются элементарными исходами) изучается также сходимость распределений вероятностей на числовой прямой или на многомерных пространствах. Каким условиям должна удовлетворять сходимость распределений вероятностей?

Первая мысль, которая кажется очень разумной — сходимость распределений $\mathbb{P}_n \rightarrow \mathbb{P}$ должна означать, что

$$\mathbb{P}_n(B) \rightarrow \mathbb{P}(B) \quad (?)$$

для любого 'хорошего' (хорошо бы, любого борелевского) множества на числовой прямой. Чтобы проверить эту идею, рассмотрим крайний случай — \mathbb{P}_n — распределения случайных величин ξ_n , чтобы все было проще, будем считать их постоянными, например, равными $1/n$. Имеем: $1/n \rightarrow 0 = \xi$. Будет разумно, если и $\mathbb{P}_{\xi_n} \rightarrow \mathbb{P}_\xi$. Но если мы выберем очень простое множество $A = \{0\}$, мы будем иметь

$$\mathbb{P}_{\xi_n}(A) = 0, \mathbb{P}_\xi(A) = 1.$$

Итак, условие (?) оказывается слишком жестким. Мы введем другую значительно более слабую сходимость, которая так и называется 'слабая сходимость' вероятностных мер. Эта сходимость тесно связана со сходимостью по вероятности.

Определение 1. Последовательность \mathbb{P}_n вероятностных распределений на борелевской числовой прямой \mathbf{R} слабо сходится к \mathbb{P} (мы пишем

$$\mathbb{P}_n \xrightarrow{\text{сл}} \mathbb{P},$$

если для любой непрерывной ограниченной функции f на \mathbf{R} имеет место сходимость интегралов

$$\int_{\mathbf{R}} f(x) d\mathbb{P}_n(x) \rightarrow \int_{\mathbf{R}} f(x) d\mathbb{P}(x). \quad (1)$$

Итак, мы ввели слабую сходимость и обещаем, что она будет подходящей для наших целей. Мы покажем, что из сходимости

$$\xi_n \xrightarrow{\mathbb{P}} \xi$$

следует

$$\mathbb{P}_{\xi_n} \xrightarrow{\text{сл}} \mathbb{P}_\xi.$$

Кроме того, будет показана возможность другого определения слабой сходимости — а именно, на языке функций распределения F_{ξ_n} и F_ξ .

В терминах слабой сходимости в теории вероятностей среди множества всех распределений на числовой прямой выделяются классы особенно важных распределений, которые являются пределами распределений встречающихся в приложениях последовательностей случайных

величин. Одна из наших целей — доказательство теорем о сходимости. Для этого придется разработать новый метод представления распределений случайных величин. В теории вероятностей эти представления называются характеристическими функциями распределений или случайных величин (не путать с характеристическими функциями множеств (событий), которые в литературе по теории вероятностей предпочитают называть индикаторами событий). В классическом анализе такого рода представления называются преобразованиями Фурье функций.

2. Слабая сходимость на языке функций распределений

Теорема. Пусть задана последовательность \mathbb{P}_n вероятностных распределений на числовой прямой \mathbf{R} и одно распределение \mathbb{P} . Напомним, что каждое вероятностное распределение \mathbb{P} на числовой прямой задает функцию распределения

$$F(x) = \mathbb{P}(-\infty, x)$$

и, в свою очередь, однозначно задается этой функцией распределения. Мы также обозначим через F_n функции распределения, определяемые вероятностными распределениями \mathbb{P}_n .

Имеем: \mathbb{P}_n слабо сходится к \mathbb{P} тогда и только тогда, когда

$$F_n(x) \rightarrow F(x) \text{ для любой } x - \text{точки непрерывности функции } F. \quad (2)$$

Комментарий. В этой теореме условие сходимости накладывается на вероятности некоторых хороших множеств A вида $(-\infty, x)$, но не для всех x . Это естественно, если x — точка разрыва функции F , то задача формулировки соответствующего условия осложняется — что считать значением $F(x)$ — предел справа $F(x+0)$ или предел слева $F(x-0)$.

Замечание. В определении в формулировке теоремы 1 мы упоминаем точки непрерывности функции распределения F . Но существуют ли такие точки? Легко показать, что число точек разрыва не более чем счетно (оно является объединением конечных множеств точек разрыва не более $1/2$ (таких точек не больше 2), $1/3, \dots, 1/n, \dots$). А мощность любого интервала в \mathbf{R} несчетна. Таким образом, в любой окрестности любой точки в \mathbf{R} (как справа, так и слева) имеются точки непрерывности функции F .

Доказательство. Как доказать, что (2) влечет (1)? Итак, нужно доказать сходимость

$$\int_{\mathbf{R}} f(x) d\mathbb{P}_n(x) \rightarrow \int_{\mathbf{R}} f(x) d\mathbb{P}(x).$$

Идея доказательства состоит в сведении доказательства сходимости интегралов к доказательству сходимости интегральных сумм.

Несколько раз используется следующее простое утверждение — аналог общеизвестного правила двух милиционеров:

Лемма 1. $x_n \rightarrow x$ тогда и только тогда, когда для любого $\varepsilon > 0$ существует последовательность $y_n \rightarrow y$ такая, что $|x_n - y_n| < \varepsilon$, начиная с некоторого n_0 , и $|x - y| < \varepsilon$.

Лемма 2. Другой вариант правила двух милиционеров — наличие двух ε -милиционеров для любого $\varepsilon > 0$:

$$y_n \leq x_n \leq z_n, y_n \rightarrow y, z_n \rightarrow z, |x - y| < \varepsilon, |x - z| < \varepsilon.$$

Доказательства этих утверждений очевидны. Мы приведем доказательство первого факта, но только для справок. Учить доказательство этого факта не нужно, (но надо самим уметь это доказывать). Итак, выберем $\varepsilon = \frac{\varepsilon}{3}$. Для $n \geq n_0$ имеем $|x_n - y_n| < \frac{\varepsilon}{3}$. Кроме того, $|x - y| < \frac{\varepsilon}{3}$. По определению предела для $n \geq n_1$ имеем $|y_n - y| < \frac{\varepsilon}{3}$. В итоге,

$$|x_n - x| \leq |x_n - y_n| + |y_n - y| + |y - x| < \frac{3\varepsilon}{3} \text{ для } n \geq \max\{n_0, n_1\}.$$

Лемма 1 позволяет сводить задачу доказательства сходимости одной последовательности к задаче доказательства сходимости другой последовательности.

Вернемся к доказательству теоремы.

Напомним, что приближение интеграла Римана интегральной суммой оказалось возможным благодаря равномерной непрерывности непрерывной функции на отрезке. В нашем случае мы интегрируем не по отрезку, а по бесконечной числовой прямой.

Итак, первая цель — заменить интегралы в (1) по числовой прямой интегралами по отрезку. Делается это следующим образом: мы подбираем такие $a < b$, что

$$1 - F(b) < \varepsilon/4C, \quad F(a) < \varepsilon/4C, \quad (*)$$

где $= \sup\{|f(x)|x \in \mathbf{R}\}$. Сдвигая, если понадобится, a влево, а b — вправо, мы добьемся того, что a и b будут точки непрерывности F .

В нашем критерии сходимости мы берем

$$\begin{aligned} x_n &= \int_{\mathbf{R}} f(x) d\mathbb{P}_n(x), \quad y_n = \int_{[a,b]} f(x) d\mathbb{P}_n(x), \\ x &= \int_{\mathbf{R}} f(x) d\mathbb{P}(x), \quad y = \int_{[a,b]} f(x) d\mathbb{P}(x). \end{aligned}$$

Действительно, из условия теоремы 1 следует, что для $n > n_0$

$$1 - F_n(b) < \varepsilon/2C, \quad F_n(a) < \varepsilon/2C,$$

$$\begin{aligned} |x_n - y_n| &= \left| \int_{\mathbf{R}} f(x) d\mathbb{P}_n(x) - \int_{[a,b]} f(x) d\mathbb{P}_n(x) \right| \leq \int_{(-\infty, a) \cup [b, \infty)} |f(x)| d\mathbb{P}_n(x) \leq \\ &\leq C(\mathbb{P}(-\infty, a) + \mathbb{P}[b, \infty)) = C\{(1 - F_n(b)) + F_n(a)\} \leq \varepsilon, \end{aligned}$$

точно так же

$$|x - y| \leq \varepsilon.$$

Теперь мы должны доказать $y_n \rightarrow y$. Мы опять используем лемму 1. Для доказательства сходимости (y_n) мы заменяем интегралы интегральными суммами. Используя равномерную непрерывность f на $[a, b]$, получаем такое разбиение $a = u_0 < u_1 < \dots < u_r = b$, что функция f на любом из отрезков разбиения меняется меньше чем на ε , но при этом мы накладываем на точки u_i дополнительное условие, они должны быть точками непрерывности функции F . Теперь нам нужно доказать

$$x_n = \int_{[a,b]} f(x) d\mathbb{P}_n(x) \rightarrow x = \int_{[a,b]} f(x) d\mathbb{P}(x).$$

Роль y_n играют интегральные суммы

$$\sum_{i=0}^{r-1} f(v_i)[F_n(u_{i+1}) - F_n(u_i)],$$

где $v_i \in (u_i, u_{i+1})$ фиксированы. Дальнейшее ясно:

$$\sum_{i=0}^{r-1} f(v_i)[F_n(u_{i+1}) - F_n(u_i)] \rightarrow \sum_{i=0}^{r-1} f(v_i)[F(u_{i+1}) - F(u_i)].$$

Но нужно еще проверить неравенства, связывающие x_n , x и y_n , y в этой части доказательства. Обозначим

$$\tilde{f}(x) = \sum_{i=0}^{r-1} f(v_i) I_{[u_i, u_{i+1})}, \quad x \in [a, b].$$

Из выбора разбиения следует, что $|f - \tilde{f}| < \varepsilon$. Поэтому

$$\begin{aligned} & \left| \int_{[a,b)} f(x) d\mathbb{P}(x) - \sum_{i=0}^{r-1} f(v_i)[F(u_{i+1}) - F(u_i)] \right| = \\ & = \sum_{i=0}^{r-1} \int_{[u_{i+1}, u_{i+1})} |f(x) - \tilde{f}(x)| d\mathbb{P}(x) \leq \varepsilon \sum_{i=0}^{r-1} [F(u_{i+1}) - F(u_i)] \leq \varepsilon \cdot 1. \end{aligned}$$

То же верно для интегралов по $\mathbb{P}_n(x)$.

Импликация (1) \Rightarrow (2) доказывается путем подбора функций f . Теперь нужно для фиксированной точки x непрерывности функции F доказать сходимость $F_n(x) \rightarrow F(x)$. Так как x — точка непрерывности, существует окрестность $[x - \delta, x + \delta]$ точки x , в которой функция F меняется не больше чем на ε . Заметим, что ввиду монотонности F для этого достаточно

$$F(x + \delta) - F(x) \leq \varepsilon, \quad F(x) - F(x - \delta) \leq \varepsilon$$

По этой окрестности мы подберем две функции — f_1 и f_2 :

$$f_1(y) = 1 \text{ при } y \in (-\infty, x - \delta], \quad f_1(y) = 0 \text{ при } y \in [x, \infty),$$

а между точками $x - \delta$ и x функция f_1 убывает линейно.

Аналогично строится функция f_2 :

$$f_2(y) = 1 \text{ при } y \in (-\infty, x], \quad f_2(y) = 0 \text{ при } y \in [x + \delta, \infty),$$

а между точками x и $x + \delta$ функция f_2 убывает линейно.

Легко проверяется, что последовательности интегралов

$$\int_{\mathbf{R}} f_1(x) d\mathbb{P}_n(x), \quad \int_{\mathbf{R}} f_2(x) d\mathbb{P}_n(x)$$

являются ε -милиционерами для последовательности $F_n(x)$ и числа $F(x)$.

Действительно, очевидно, что

$$I_{(-\infty, x-\delta)} \leq f_1 \leq I_{(-\infty, x)} \leq f_2 \leq I_{(-\infty, x+\delta)}.$$

Поэтому

$$\int_{\mathbf{R}} f_1(y) d\mathbb{P}_n(y) \leq F_n(x) \leq \int_{\mathbf{R}} f_2(y) d\mathbb{P}_n(y) \text{ для всех } n,$$

$$\int_{\mathbf{R}} f_1(y) d\mathbb{P}_n(y) \rightarrow \int_{\mathbf{R}} f_1(y) d\mathbb{P},$$

$$\int_{\mathbf{R}} f_2(y) d\mathbb{P}_n(y) \rightarrow \int_{\mathbf{R}} f_2(y) d\mathbb{P}(y),$$

$$F(x) - \varepsilon \leq F(x - \delta) \leq \int_{\mathbf{R}} f_1(y) d\mathbb{P}(y) \leq F(x) \leq \int_{\mathbf{R}} f_2(y) d\mathbb{P}(y) \leq F(x + \delta) \leq F(x) + \varepsilon.$$

Упражнение. Пусть \mathbb{P}_n — нормальные распределения с параметрами $N(m_n, \sigma_n)$, \mathbb{P} — нормальное распределение с параметрами $N(m, \sigma)$. Дано, что $m_n \rightarrow m$, $\sigma_n \rightarrow \sigma$. Доказать

$$\mathbb{P}_n \xrightarrow{\text{сл}} \mathbb{P}.$$

Если $t_n = 0$, $\sigma_n \rightarrow 0$, то \mathbb{P}_n слабо сходится к распределению, сосредоточенному в точке 0 (такое распределение часто называется функцией Дирака δ).

3. Теорема. Если ξ_n , ξ – случайные величины,

$$\xi_n \xrightarrow{\mathbb{P}} \xi,$$

то

$$\mathbb{P}_{\xi_n} \xrightarrow{\text{сл}} \mathbb{P}_\xi.$$

Доказательство использует формулу (2) в качестве определения слабой сходимости. Надо доказать:

$$F_{\xi_n}(x) \rightarrow F_\xi(x) \text{ для любой точки } x \text{ непрерывности } F_\xi.$$

Мы опять фиксируем такое δ , что F_ξ изменяется в окрестности $[x - \delta, x + \delta]$ не более чем на ε , кроме того, фиксируем n_0 , начиная с которого

$$\mathbb{P}\{|\xi_n - \xi| \geq \delta\} < \varepsilon.$$

Непосредственно проверяются два включения

$$\{\xi < x - \delta\} \subset \{\xi_n < x\} \cup \{|\xi_n - \xi| \geq \delta\}, \quad (*)$$

$$\{\xi_n < x\} \subset \{\xi < x + \delta\} \cup \{|\xi_n - \xi| \geq \delta\}. \quad (**)$$

Проверим, например, первое включение: если

$$\xi(\omega) < x - \delta, \quad \xi_n(\omega) \geq x, \quad \text{то } |\xi_n(\omega) - \xi(\omega)| \geq \delta.$$

Далее мы видим: (*) влечет

$$F_\xi(x) - \varepsilon \leq F_\xi(x - \delta) \leq F_{\xi_n}(x) + \mathbb{P}\{|\xi_n - \xi| \geq \delta\} \leq F_{\xi_n}(x) + \varepsilon,$$

(**) влечет

$$F_{\xi_n}(x) \leq F_\xi(x + \delta) + \mathbb{P}\{|\xi_n - \xi| \geq \delta\} \leq F_\xi(x) + 2\varepsilon.$$

Вместе эти неравенства ввиду произвольности ε означают нужную сходимость.

Пример. Очевидно, что из слабой сходимости распределений случайных величин F_{ξ_n} к F_ξ не следует сходимость по вероятности случайных величин ξ_n к ξ . Приведем простой пример:

$$\Omega = \{0, 1\}, \mathbb{P}\{0\} = \frac{1}{2}, \mathbb{P}\{1\} = \frac{1}{2}, \xi_n(1) = 0, \xi_n(0) = 1, \xi(0) = 0, \xi(1) = 1.$$

Имеем:

$$\mathbb{P}\{|\xi_n - \xi| = 1\} = 1,$$

в то же время

$$\mathbb{P}_{\xi_n} = \mathbb{P}_\xi,$$

и тем более имеет место сходимость

$$\mathbb{P}_{\xi_n} \xrightarrow{\text{сл}} \mathbb{P}_\xi.$$

Этот пример основан на наличии случайной величины η , имеющих то же распределение, что и ξ , но не равных ξ . Более того, любая такая пара $\{\xi, \eta\}$ дает контрпример $\eta, \eta, \eta, \dots \not\rightarrow \xi$. Возможность выбора пары зависит также от вероятностного пространства. Например, если вероятность одной точки равна $2/3$, а другой – $1/3$, такой пары не существует. Но если $\xi = \text{Const}$, то есть постоянная случайная величина, то такой контрпример оказывается невозможным. В этом случае верно обратное утверждение.

Обозначения. Слабая сходимость — сходимость распределений, а не случайных величин. Однако нам будет удобно следующее обозначение:

$$\xi_n \xrightarrow{\mathcal{D}} \xi \iff \mathbb{P}_{\xi_n} \xrightarrow{\text{с.п.}} \mathbb{P}_\xi,$$

при этом мы будем говорить, что ξ_n сходится к ξ по распределению (in distribution). Точно так же мы будем использовать обозначение

$$\xi \stackrel{\mathcal{D}}{=} \eta \iff \mathbb{P}_\xi = \mathbb{P}_\eta,$$

и будем говорить, что случайные величины ξ и η совпадают по распределению.

4. Теорема. Если ξ_n, ξ — случайные величины, $\xi = c$,

$$\mathbb{P}_{\xi_n} \xrightarrow{\text{с.п.}} \mathbb{P}_\xi,$$

то

$$\xi_n \xrightarrow{\mathbb{P}} \xi.$$

Доказательство. В этой ситуации все точки $x \neq c$ являются точками непрерывности функции распределения F_ξ . Пусть даны $\delta > 0$ и $\varepsilon > 0$, возьмем $x = c + \delta$ и $x = c - \delta$. Выберем по ним такое n_0 , что для всех $n > n_0$ имеют место неравенства

$$|F_{\xi_n}(+ \delta) - F_\xi(+ \delta)| < \frac{\varepsilon}{2}, |F_{\xi_n}(- \delta) - F_\xi(- \delta)| < \frac{\varepsilon}{2}.$$

Так как

$$F_\xi(- \delta) = 0, F_\xi(+ \delta) = 1,$$

это означает

$$F_{\xi_n}(+ \delta) > 1 - \frac{\varepsilon}{2}, F_{\xi_n}(- \delta) < \frac{\varepsilon}{2}.$$

Рассмотрим

$$\begin{aligned} \mathbb{P}\{|\xi_n - \xi| > \delta\} &= \mathbb{P}\{\xi_n > c + \delta\} + \mathbb{P}\{\xi_n < c - \delta\} \leq \\ &\leq (1 - F_{\xi_n}(+ \delta)) + F_{\xi_n}(- \delta) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2}. \end{aligned}$$

Мы получили сходимость по вероятности.

Комментарий. Итак, имеется связь между слабой сходимостью и сходимостью по вероятности. Но не нужно путать эти два понятия. Сходимость по вероятности относится к случайным величинам, то есть к функциям, заданным на множестве элементарных исходов. Слабая сходимость относится к распределениям, то есть к вероятностям, заданным на борелевских множествах вещественной прямой (которая, кстати, состоит из чисел, а не из элементарных исходов). Добавим к этому то, что иногда мы создаем вероятностные модели, в которых элементарные исходы — числа в \mathbf{R} или векторы в \mathbf{R}^n , события — борелевские множества, а вероятность — распределение некоторой случайной величины или некоторого случайного вектора.

Упражнение 1. Возьмем в качестве вероятностного пространства $(0, 1)$ с σ -алгеброй борелевских множеств как σ -алгеброй событий и вероятностью, равной мере Лебега. Покажите, что для любой строго возрастающей непрерывной функции распределения F функция F^{-1} — случайная величина на $(0, 1)$ с функцией распределения F . Используя эту конструкцию, придумайте две разные случайные величины на $(0, 1)$ с нормальным распределением со средним 0 и дисперсией 1.

Упражнение 2. Проверьте целесообразность введения в понятие слабой сходимости на языке интегралов условий ограниченности и непрерывности произвольной функции.

а) Покажите, что если измеримая функция f на \mathbf{R} неограничена, то существует такое одномерное распределение \mathbb{P} , что $\int f(x)d\mathbb{P}(x)$ расходится. *Намек.* Распределение возьмите дискретным, тогда интеграл превратится в ряд, нужно, чтобы этот ряд сходился к ∞ или $-\infty$. Из

этого упражнения следует, что определение слабой сходимости оказывается неуниверсальным. Поэтому для искомой вероятности даже утверждение

$$\mathbb{P}, \mathbb{P}, \mathbb{P}, \dots \xrightarrow{\text{сл}} \mathbb{P})$$

не имеет смысла.

b) Пусть $f(x) = I_{(-\infty, 0]}$, $\mathbb{P}_n\{1/n\} = 1$ — распределения констант, сходящихся к нулю, $\mathbb{P}\{0\} = 1$. Тогда $\int_{\mathbf{R}} f(x)d\mathbb{P}_n(x) = 0$, $\int_{\mathbf{R}} f(x)d\mathbb{P}(x) = 1$.

Упражнение 18. Если $f : \mathbf{R} \rightarrow \mathbf{R}$ — непрерывная функция,

$$\mathbb{P}_{\xi_n} \xrightarrow{\text{сл}} \mathbb{P}_{\xi},$$

то

$$\mathbb{P}_{f(\xi_n)} \xrightarrow{\text{сл}} \mathbb{P}_{f(\xi)}.$$

§13. Характеристическая функция

Для изучения слабой сходимости распределений эффективен аппарат *характеристических функций* (преобразований Фурье распределений).

1. Характеристическая функция.

Определение. *Характеристической функцией* φ вероятностного распределения \mathbb{P} на числовой прямой называется функция

$$\varphi(t) = \int_{\mathbf{R}} e^{itx} d\mathbb{P}(x).$$

Если нет возможности путаницы, мы будем обозначать характеристическую функцию $\varphi_{\mathbb{P}}$ или φ_F , если распределение задано функцией распределения (или φ_{ξ} , если наше распределение является распределением случайной величины ξ), в этом случае мы можем ее определить соотношением

$$\varphi_{\xi}(t) = \mathbb{E}(e^{it\xi}),$$

то же соотношение в интегральных обозначениях записывается так:

$$\varphi_{\xi}(t) = \int_{\Omega} e^{it\xi} d\mathbb{P}.$$

Если использовать интеграл Римана – Стильтьеса, получается

$$\varphi_{\xi}(t) = \int_{\mathbf{R}} e^{itx} dF_{\xi}(x).$$

Если имеется функция плотности распределения, мы имеем несобственный интеграл:

$$\varphi_{\xi}(t) = \int_{\mathbf{R}} e^{itx} p_{\xi}(x) dx.$$

Наконец, в случае дискретного распределения, мы имеем сумму:

$$\varphi_{\xi}(t) = \sum_n e^{ita_n} \mathbb{P}\{\xi = a_n\}.$$

2. Свойства характеристических функций

1. $\varphi(0) = 1; \varphi(-t) = \bar{\varphi}(t)$.

2. φ — непрерывная и даже равномерно непрерывная функция;

Непрерывность немедленно следует из теоремы Лебега и определения непрерывности на языке последовательностей. Пусть $t_n \rightarrow t$, тогда $|e^{it_n x}| = 1 \leq 1$ и $e^{it_n x} \rightarrow e^{it x}$ для всех x , а функция $f(x) = 1$ интегрируется по вероятностной (!) мере $dF_\xi(x)$. По теореме Лебега

$$\int_{\mathbf{R}} e^{it_n x} dF_\xi(x) \rightarrow \int_{\mathbf{R}} e^{it x} dF_\xi(x).$$

Равномерная непрерывность характеристической функции доказывается тощше. Сначала по данному $\varepsilon > 0$ (используя сходимость функций распределения на $\pm\infty$) мы выбираем такие $a, b \in \mathbf{R}$, что

$$\int_{[a,b]} dF(x) > 1 - \frac{\varepsilon}{4}.$$

Так как функция $u \rightarrow e^{iu}$ непрерывна в нуле, то для некоторого $\delta_1 > 0$ верно $|1 - e^{iu}| < \varepsilon/2$ для всех $|u| < \delta_1$. Далее выберем $0 < \delta < \delta_1 / \max |a|, |b|$. Тогда из $|t| < \delta$, $x \in [a, b]$ следует $|1 - e^{itx}| < \varepsilon/2$. Если теперь $|t - t'| < \delta$, то

$$\begin{aligned} |\varphi_F(t) - \varphi_F(t')| &= \left| \int_{\mathbf{R}} e^{itx} - e^{it'x} dF(x) \right| \leq \int_{\mathbf{R}} |e^{itx} - e^{it'x}| dF(x) = \\ &= \int_{[a,b]} + \int_{\mathbf{R} \setminus [a,b]} \leq \int_{[a,b]} \varepsilon/2 dF(x) + \int_{\mathbf{R} \setminus [a,b]} 2dF(x) \leq \varepsilon. \end{aligned}$$

3. Если случайные величины ξ и η независимы, то

$$\varphi_{\xi+\eta}(t) = \varphi_\xi(t)\varphi_\eta(t);$$

4.

$$\varphi_{a\xi+b}(t) = e^{itb}\varphi_\xi(at);$$

5. Если случайная величина ξ имеет среднее, то дифференцирование под знаком интеграла дает нам

$$\varphi'_\xi(0) = i\mathbb{E}(\xi),$$

поэтому

$$\varphi_\xi(t) = 1 + i\mathbb{E}(\xi)t + o(t).$$

Если случайная величина ξ имеет дисперсию, то

$$\varphi''_\xi(0) = -\mathbb{E}(\xi^2),$$

$$\varphi_\xi(t) = 1 + i\mathbb{E}(\xi)t - \frac{1}{2}\mathbb{E}(\xi^2)t^2 + o(t^2).$$

6. **Теорема единственности.** Характеристическая функция φ_F однозначно определяет функцию распределения F (а следовательно, и распределение).

7. **Теорема непрерывности.** Последовательность распределений сходится слабо тогда и только тогда, когда последовательность их функций распределения сходится поточечно.

Комментарий. Свойства 1-4 проверяются непосредственной проверкой. Причем свойство 3 — очевидное следствие свойства среднего — среднее произведение независимых случайных величин равно произведению средних. Свойство 5 получается разложением Тейлора характеристической функции с одновременным дифференцированием под знаком среднего. Прежде

чем доказать свойство 6, выводится формула обращения, которая позволяет выражать через характеристическую функцию разность значений функции распределения в двух точках непрерывности функции распределения. Доказательство теоремы непрерывности сложно и использует теоремы о компактности некоторого расширения множества функций распределений, критерий слабой компактности в пространстве функций распределения, а также одно неравенство, позволяющее проверить этот критерий в терминах поведения функций распределения в окрестности нуля.

3. Как вычислять характеристическую функцию?

Если ξ имеет дискретное распределение ($\sum_n a_n \mathbb{P}\{\xi = a_n\} = 1$), то

$$\varphi_\xi(t) = \sum_n e^{ia_n t} \mathbb{P}\{\xi = a_n\}.$$

Если ξ имеет непрерывный тип распределения с функцией плотности $p_\xi(x)$, то

$$\varphi_\xi(t) = \int_{\mathbf{R}} e^{itx} p_\xi(x).$$

4. Примеры.

1. Характеристическая функция распределения Пуассона.

$$\mathbb{P}\{k\} = \frac{\lambda^k}{k!} e^{-\lambda} \varphi(t) = e^{\lambda(e^{it}-1)}.$$

2. Характеристическая функция нормального распределения.

Сначала рассмотрим параметры $N(0, 1)$, $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Тогда

$$\varphi(t) = \int_{\mathbf{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2} + itx} dx = e^{-\frac{t^2}{2}}.$$

Этот интеграл можно найти в Демидовиче. Но можно вычислить и самостоятельно, представив его как интеграл по прямой \mathbf{R} на комплексной плоскости \mathbf{C} и используя известную теорему о независимости от пути интеграла от аналитической функции.

$$\begin{aligned} \varphi(t) &= \int_{\mathbf{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2} + itz} dz = e^{-\frac{t^2}{2}} \int_{\mathbf{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-it)^2}{2}} dz = \\ &= e^{-\frac{t^2}{2}} \int_{\mathbf{R}-it} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = e^{-\frac{t^2}{2}} \int_{\mathbf{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = e^{-\frac{t^2}{2}}. \end{aligned}$$

Общий случай сводится к предыдущему: нормальное распределение с параметрами t и σ имеет случайная величина $\sigma\xi + t$, где ξ имеет параметры 0, 1. По свойству 4

$$\varphi_{\sigma\xi+t}(t) = e^{itm} \varphi_\xi(\sigma t) = e^{itm - \frac{1}{2}\sigma^2 t^2}.$$

Элементарным следствием из этих выкладок и теоремы единственности является устойчивость распределений Пуассона и нормального: сумма независимых случайных величин с распределением Пуассона имеет распределение Пуассона, сумма независимых случайных величин с нормальным распределением имеет нормальное распределение.

3. Характеристическая функция распределения Коши.

Упражнение. Докажите, что

$$\int_{-\infty}^{\infty} e^{itx} \frac{1}{\pi(1+x^2)} dx = e^{-|t|}.$$

Покажите, что если ξ_1, \dots, ξ_n независимы и имеют распределение Коши, то

$$\frac{\xi_1 + \dots + \xi_n}{n}$$

имеет распределение Коши.

5. Центральная предельная теорема. Пусть (ξ_n) — последовательность одинаково распределенных независимых случайных величин со средним и дисперсией, обозначим

$$\mathbb{E}(\xi_n) = m, \mathbb{D}(\xi_n) = \sigma^2.$$

Тогда последовательность распределений центрированных и нормированных сумм

$$\frac{\xi_1 + \xi_2 + \dots + \xi_n - nm}{\sigma\sqrt{n}}$$

слабо сходится к нормальному распределению с параметрами $N(0, 1)$.

Доказательство основано на разложении (свойство 5), теоремах единственности и непрерывности, свойствах 3 и 4.

Запишем разложение 5 для характеристической функции случайной величины $\frac{\xi_i - m}{\sigma}$. У этой случайной величины среднее равно нулю, а дисперсия — 1. Поэтому

$$\varphi_{\frac{\xi_i - m}{\sigma}}(t) = 1 + 0 \cdot it - \frac{t^2}{2} + o(t^2).$$

По свойству 3 имеем:

$$\varphi_{\frac{\xi_1 + \xi_2 + \dots + \xi_n - nm}{\sigma\sqrt{n}}} = \left(1 - \frac{t^2}{2} + o(t^2)\right)^n.$$

По свойству 4

$$\varphi_{\frac{\xi_1 + \xi_2 + \dots + \xi_n - nm}{\sigma\sqrt{n}}} = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{2n}\right)\right)^n \rightarrow e^{-\frac{t^2}{2}}.$$

По теореме непрерывности мы имеем сходимости, а по теореме единственности идентифицируем распределение с данной характеристической функцией.

Упражнения 19. а. Докажите, что слабый предел нормальных распределений — нормальное распределение (здесь мы считаем распределение константы вырожденным нормальным распределением).

б. Докажите, что слабый предел распределений Пуассона — распределение Пуассона (здесь мы считаем распределение константы вырожденным распределением Пуассона).

с. Докажите, что слабый предел равномерных распределений — равномерное распределение (здесь мы считаем распределение константы вырожденным равномерным распределением).

д. Может ли предел по вероятности случайных величин с нормальным распределением иметь распределение, не являющееся нормальным? (Как и выше, здесь мы считаем распределение константы вырожденным нормальным распределением).

е. Даны распределения \mathbb{P}_n , задаваемые равенствами $\mathbb{P}_n\{k/n\} = 1/n$ для $k = 1, 2, \dots, m$. Доказать, что распределения \mathbb{P}_n слабо сходятся к равномерному распределению на отрезке $[0, 1]$.

Важнейшие свойства характеристических функций

Теорема единственности

Формула обращения

Пусть a и b — две точки непрерывности функции распределения F . Тогда

$$F(b) - F(a) = \lim_{A \rightarrow \infty} \frac{1}{2\pi} \int_{-A}^A \frac{e^{-ita} - e^{-itb}}{it} \varphi_F(t) dt. \quad (1)$$

Интеграл в правой части является интегралом в смысле главного значения, если заменить предел на интеграл по всей числовой прямой, то возникнет неочевидный вопрос о сходимости интеграла, во многих случаях, например, для случая $\varphi_F(t) = \cos tb = \frac{e^{itb} + e^{-itb}}{2}$ интеграл по $t \in \mathbf{R}$ в обычном смысле расходится. В математическом анализе имеются понятия преобразования Фурье и обратного преобразования Фурье, которое с точностью до нормирующего множителя выглядит так:

$$p(x) = \frac{1}{2\pi} \int_{\mathbf{R}} e^{-itx} \varphi_F(t) dt,$$

где p — функция плотности для F , которая должна удовлетворять некоторым дополнительным условиям. Формула (1) получается из предыдущей интегрированием от a до b , но выполняется при значительно более общих условиях. Плотности и, тем более, 'хорошей плотности' у распределения может и не быть.

Для доказательства (1) мы заменим φ_F на свое определение и получим двойной интеграл:

$$\lim_{A \rightarrow \infty} \frac{1}{2\pi} \int_{-A}^A \frac{e^{-ita} - e^{-itb}}{it} dt \int_{\mathbf{R}} e^{itx} dF(x).$$

Нетрудно проверить, что подынтегральная функция ограничена на произведении $[-A, A] \times \mathbf{R}$, проблемы возможны лишь в точке $t = 0$, но там предел $\frac{e^{-ita} - e^{-itb}}{it}$ вычисляется по правилу Лопиталя, т.е. существует. Обе меры, по которым мы интегрируем, и на $[-A, A]$ и на \mathbf{R} , конечны. Мы вправе применить теорему Фубини и записать все следующим образом:

$$\lim_{A \rightarrow \infty} \frac{1}{2\pi} \int_{\mathbf{R}} dF(x) \int_{-A}^A \frac{e^{-ita} - e^{-itb}}{it} e^{itx} dt. \quad (2)$$

Мы вводим дополнительное обозначение и сводим внутренний интеграл к следующему:

$$I_A(x) = \int_{-A}^A \frac{e^{-ita} - e^{-itb}}{it} e^{itx} dt = 2 \int_0^A \frac{\sin((x-a)t)}{t} - \frac{\sin((x-b)t)}{t} dt.$$

Теперь мы хотим перенести предел в (2), то есть доказать

$$\lim_{A \rightarrow \infty} \frac{1}{2\pi} \int_{\mathbf{R}} I_A(x) dF(x) = \frac{1}{2\pi} \int_{\mathbf{R}} \lim_{A \rightarrow \infty} I_A(x) dF(x),$$

для этого мы воспользуемся теоремой Лебега. Чтобы быть вправе использовать теорему Лебега (она формулируется для последовательностей), мы будем доказывать предельное соотношение на языке последовательностей. Итак, нам достаточно доказать

- i) сходимость $\lim_{A_n \rightarrow \infty} I_{A_n}(x)$ для всех x ,
- ii) ограниченность функции $I_A(x)$ некоторой константой C (функция $g(x) = C$ очевидно интегрируема по мере dF).
- i) следует из известной вам сходимости интеграла

$$\int_0^\infty \frac{\sin \alpha t}{t} dx,$$

к которому сходятся оба интеграла в определении $I_A(x)$, с $\alpha = (x - a)$ и с $\alpha = (x - b)$.

ii) следует из ограниченности всех интегралов вида $\int_0^B \frac{\sin t}{t} dt$, которая следует из сходимости интеграла $\int_0^\infty \frac{\sin t}{t} dt$ (ограниченность в некоторой окрестности ∞ , то есть для всех $B \in (B_1, \infty)$), и из непрерывности функции $\int_0^B \frac{\sin t}{t} dt$ на отрезке $[0, B_1]$. Итак,

$$\left| \int_0^B \frac{\sin t}{t} dt \right| < C, \forall B,$$

в то же время

$$\int_0^A \frac{\sin((x-a)t)}{t} dt = \operatorname{sgn}(A(x-a)) \int_0^{|A(x-a)|} \frac{\sin t}{t} dt.$$

Теперь мы можем завершить доказательство формулы обращения. Обозначим

$$I(x) = \int_0^\infty \frac{\sin((x-a)t)}{t} - \frac{\sin((x-b)t)}{t} dt =$$

$$= \begin{cases} 0, & \text{если } x > a, x > b; \\ \pi, & \text{если } a < x < b; \\ 0, & \text{если } x < a, x < b; \\ \pi/2, & \text{если } x = b; \\ -\pi/2, & \text{если } x = b. \end{cases}$$

Заметим, что значения I в точках a и b не существенны, так как меры этих точек равны нулю в силу непрерывности F . Пользуясь этим, мы подправим подинтегральную функцию в этих точках. Итак,

$$\lim_{A \rightarrow \infty} \frac{1}{2\pi} \int_{-A}^A \frac{e^{-ita} - e^{-itb}}{it} \varphi_F(t) dt =$$

$$= \int_{\mathbb{R}} \frac{1}{2\pi} 2I(x) dF(x) = \int_{\mathbb{R}} I_{[a,b]}(x) dF(x) = F(b) - F(a).$$

Доказательство теоремы единственности. Пусть F и G — две различные функции распределения, имеющие одинаковую характеристическую функцию. Согласно формуле обращения, для любых чисел x и y , которые являются точками непрерывности обеих функций — F и G ,

$$F(y) - F(x) = G(y) - G(x).$$

Лемма. *Множество точек разрыва у любой функции распределения F не более чем счетное.*

Доказательство очевидно, так как

$$\{\text{множество всех точек разрыва } F\} = \bigcup_n \{\text{множество всех точек разрыва } F \text{ больше } 1/n\},$$

каждое множество в правой части конечно (может содержать самое большее $n - 1$ точек). объединение счетного числа конечных множеств счетно.

Следствия. *В любом открытом интервале содержится бесконечное число точек непрерывности функций распределения F и G .*

Следствие влечет, что существует последовательность точек непрерывности (x_n) функций F и G , сходящаяся к $-\infty$, и для любого $y \in \mathbf{R}$ существует последовательность точек непрерывности (z_n) функций F и G , сходящаяся к y слева. Переходя к пределу, имеем:

$$F(z) = \lim_n (F(x) - F(x_n)) = \lim_n (G(x) - G(x_n)) = G(z)$$

для любой точки непрерывности z , а также

$$F(y) = \lim_n F(z_n) = \lim_n G(z_n) = G(y)$$

для любого y .

Суммирование нормальных и пуассоновских независимых случайных величин.

Из теоремы единственности следуют очень интересные свойства нормальных и пуассоновских случайных величин.

Предложение. Пусть случайные величин ξ и η независимы. 1) Если ξ имеет распределение $N(m_1, \sigma_1^2)$, а $\eta - N(m_2, \sigma_2^2)$, то $\xi + \eta$ имеет распределение $N(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2})$. 2) Если ξ имеет распределение $P(\lambda_1)$, а $\eta - P(\lambda_2)$, то $\xi + \eta$ имеет распределение $P(\lambda_1 + \lambda_2)$.

Доказательство. Представим характеристическую функцию $\xi + \eta$ как произведение, по ней восстанавливается вид распределения.

Доказательство теоремы непрерывности. В одну сторону утверждение уже доказано, по теореме об эквивалентности определений слабой сходимости на языке функций распределения и на языке интегралов в качестве непрерывной ограниченной функции мы можем взять функцию $f(x) = e^{itx}$ и из слабой сходимости F_n к F следует $\varphi_{F_n}(t) \rightarrow \varphi_F(t)$ для всех t . Мы воспользуемся этим фактом (прием Мюнхаузена), чтобы доказать обратную импликацию: $\varphi_{F_n}(t) \rightarrow \varphi_F(t)$ для всех t влечет слабую сходимость F_n к F . Но сначала надо доказать теорему о свойстве компактности множества функций распределения слабо сходящейся подпоследовательности. Такое утверждение неверно для множества (Δ) всех функций распределений, то есть функций, удовлетворяющих условиям 1)-3). Например, какую бы мы не выделяли подпоследовательность из последовательности $I_{(n, \infty)}$, где $n \rightarrow \infty$, она в каждой точке будет сходить к нулю. Таким образом, предел не будет удовлетворять условию 2). Поэтому чтобы достигнуть компактности, мы должны расширить пространство Δ и рассмотреть пространство Δ' всех функций F на \mathbf{R} , удовлетворяющих условиям 1) и 3), а условие 2) мы заменим более слабым

2') $0 \leq F(x) \leq 1$ для всех $x \in \mathbf{R}$. Возникшая в нашем контрпримере функция, тождественно равная 0, этому условию удовлетворяет.

Теорема 1. Из любой последовательности $(F_n) \subset \Delta'$ можно выделить подпоследовательность (F_{n_k}) , которая сходится к некоторой функции $F \in \Delta'$ в каждой точке непрерывности некоторой функции F .

Доказательство. Как во многих таких доказательствах, мы должны сначала построить подпоследовательность (F_{n_k}) и функцию F , а после этого доказать требуемую сходимость. Для построения мы воспользуемся знаменитой теоремой Больцано – Вейерштрасса о выделении сходящейся последовательности из ограниченной числовой последовательности (в нашем случае она находится в $[0, 1]$). Мы введем в \mathbf{R} некоторое всюду плотное счетное множество $S = (r_n)$ (например, \mathbf{Q}), потом выделим из числовой последовательности $(F_n(r_1))$ сходящуюся подпоследовательность $(F_{n_k}(r_1))$. Далее выделим подпоследовательность $(F_{n_{k_l}})$, которая имеет предел в точке r_2 , из нее выделим подпоследовательность $(F_{n_{k_{l_m}}})$, сходящуюся в точке r_2 и т. д. Наконец, используя метод выделения диагональной последовательности, мы выделим подпоследовательность $F_{n_1}, F_{n_2}, F_{n_{k_3}}, \dots$, которая сходится на всех точках r_n . Для удобства используем для этой последовательности обозначение (F_{n_k}) . Итак, $(F_{n_k}(r))$ сходится для всех $r \in S$ к некоторому числу $\alpha(r)$.

Теперь нам нужно ввести функцию F . Заметим для будущего, что функция α монотонна на S как предел последовательности монотонных функций. Итак,

$$F(y) = \sup\{\alpha(r) : r < y\}.$$

Очевидно, что значения F (как и α) лежат в $[0, 1]$. Монотонность F следует из определения супремума. Воспользуемся определением супремума на языке ε : для любого $\varepsilon > 0$ существует $r < y$ такое, что $\alpha(r) > F(y) - \varepsilon$. Имеем: для любого $x \in (r, y]$ справедливо $F(x) \geq \alpha(r) > F(y) - \varepsilon$. Это доказывает непрерывность слева в произвольной точке y .

Осталось доказать сходимость

$$F_{n_k}(x) \rightarrow F(x) \text{ для всех } x - \text{точек непрерывности } F.$$

Так как функция F монотонна, непрерывность в x эквивалентна существованию таких $y < x < z$, что $F(y) > F(z) - \varepsilon$. Выберем некоторые $r', r'' \in S$ так, что

$$y < r' < x < r'' < z.$$

Имеем по определению α :

$$\begin{aligned} F(x) - \varepsilon &\leq F(y) \leq F(r') = \lim_k F_{n_k}(r') \leq \liminf_k F_{n_k}(x) \leq \\ &\leq \limsup_k F_{n_k}(x) \leq \lim_k F_{n_k}(r'') = F(r'') \leq F(z) < F(x) + \varepsilon. \end{aligned}$$

Ввиду произвольности ε это означает

$$\lim_k F_{n_k}(x) = F(x).$$

Теперь мы изложим *план доказательства теоремы непрерывности*. Предположим, что i) $\varphi_{F_n}(t) \rightarrow \varphi_F(t)$ для всех t , но для некоторой точки x непрерывности функции распределения F сходимости $F_n(x)$ к $F(x)$ нет. Мы хотим прийти к противоречию с i). Всегда можно выделить сходящуюся подпоследовательность $F_{n_k}(x)$. Однако эта подпоследовательность может случайно оказаться сходящейся к $F(x)$, и мы никакого противоречия не получим. Поэтому мы сначала выделим из $(F_n(x))$ подпоследовательность, которая находится вне некоторой окрестности $F(x)$, а уже из нее выделим сходящуюся подпоследовательность $F_{n_k}(x)$. Имеем: $F_{n_k}(x) \rightarrow a \neq F(x)$. Для этой подпоследовательности также верно i). Согласно доказанной теореме о слабой компактности выделим из (F_n) слабо сходящуюся

$$F_{n_{k_l}} \xrightarrow{\text{сл}} G. \quad (*)$$

Согласно доказанному прямому утверждению

$$\varphi_{F_{n_{k_l}}}(t) \rightarrow \varphi_G(t) \text{ для всех } t.$$

Но так как по условию

$$\varphi_{F_{n_{k_l}}}(t) \rightarrow \varphi_F(t) \text{ для всех } t,$$

то $\varphi_F = \varphi_G$, по теореме единственности $F = G$. Но тогда согласно (*) $F_{n_k}(x) \rightarrow F(x)$, что противоречит нашему выбору подпоследовательности (F_{n_k}) .

Итак, казалось бы, теорема непрерывности доказана, но это не так. Прямое утверждение в теореме непрерывности мы вывели из эквивалентности двух определений слабой сходимости. А в доказательстве этого факта мы существенно использовали то, что предельная функция также функция распределения, принадлежит Δ , на $-\infty$ сходится к 0, а на $+\infty$ сходится к 1. Это пока ниоткуда не следует для функции G , так как теорему о слабой компактности мы доказали для Δ' , а для Δ она неверна. Итак, нам нужно показать, что предел любой подпоследовательности (F_n) принадлежит Δ . Для этого мы должны получить критерий слабой компактности

для последовательностей в Δ и доказать, что из сходимости характеристических функций (F_n) к характеристической функции распределения следует выполнение условий такого критерия.

Критерий слабой компактности в Δ . Последовательность (F_n) в Δ слабо компактна в Δ , то есть из любой ее подпоследовательности (F_{n_k}) можно выделить подпоследовательность $(F_{n_{k_l}})$, сходящуюся к элементу Δ , тогда и только тогда, когда ii) для любой $\varepsilon > 0$ существуют $N \in \mathbf{N}$ и $A > 0$ такие, что для всех $n > N$ имеет место

$$F_n(-A) < \varepsilon, F_n(+A) > 1 - \varepsilon.$$

Доказательство очевидно. По теореме о слабой компактности в Δ' существует такая подпоследовательность

$$F_{n_{k_l}} \xrightarrow{\text{с.л.}} G \in \Delta'.$$

С другой стороны по условию для точек непрерывности $a < -A$ и $b > A$ функции G выполняется

$$G(a) \leq \varepsilon, G(b) \geq 1 - \varepsilon, \quad (**)$$

$G \in \Delta$ ввиду произвольности ε .

Далее мы докажем неравенство для характеристических функций, которое позволяет доказывать оценки для $\int_{|x| \geq A} dF(x)$ через характеристическую функцию φ_F , что позволит нам проверить выполнение условия критерия слабой компактности.

Неравенство для характеристической функции.

$$\int_{|x| \geq A} dF(x) \leq \frac{1}{1 - \sin 1} \frac{A}{2} \int_{-1/A}^{1/A} (1 - \varphi_F(t)) dt.$$

Доказательство.

$$\begin{aligned} \frac{A}{2} \int_{-1/A}^{1/A} (1 - \varphi_F(t)) dt &= \frac{A}{2} \int_{-1/A}^{1/A} \int_{-\infty}^{\infty} (1 - e^{itx}) dt dF(x) = \\ &= A \int_0^{1/A} \int_{-\infty}^{\infty} (1 - \cos(tx)) dt dF(x) = \int_{-\infty}^{\infty} \frac{x/A - \sin(x/A)}{x/A} dF(x) \geq \\ &\geq (1 - \sin(1)) \int_{|x| \geq A} dF(x). \end{aligned}$$

Последнее неравенство следует из легко проверяемого неравенства

$$1 - \sin(1) \leq \frac{x/A - \sin(x/A)}{x/A}, |x| > A.$$

Последний этап доказательства теоремы непрерывности — проверка слабой компактности последовательности (F_n) в Δ .

Выберем по данному $\varepsilon > 0$ такое $A > 0$, что для всех $t \in [-1/A, 1/A]$ имело бы место $\frac{1}{1 - \sin 1} |1 - \varphi_F(t)| < \varepsilon$, тогда

$$\frac{1}{1 - \sin 1} \frac{A}{2} \int_{-1/A}^{1/A} |1 - \varphi_F(t)| dt < \varepsilon.$$

Так как $\varphi_{F_n}(t) \rightarrow \varphi_{F_n}(t)$ для всех t , по теореме Лебега мы будем иметь для всех n начиная с некоторого N также

$$\frac{1}{1 - \sin 1} \frac{A}{2} \int_{-1/A}^{1/A} |1 - \varphi_{F_n}(t)| dt < \varepsilon.$$

Согласно неравенству для характеристических функций и критерию слабой компактности, мы получили слабую компактность (F_n) в Δ , что и требовалось.

§14. Теорема Линдеберга

Используемые неравенства:

$$a) |e^{ix} - 1| \leq |x|, b) |e^{ix} - ix - 1| \leq x^2/2.$$

Первое неравенство можно рассматривать как неравенство без i на отрезке комплексной плоскости от 0 до ix . Второе неравенство получается интегрированием первого по этому отрезку. Третье неравенство следует из разложения в ряд Тейлора:

$$c) |\ln(1 + z) - z| \leq C|z|^2$$

в окрестности 0 комплексной плоскости вида $\{|z| < 1/2\}$.

Теорема. (Линдеберг) Пусть ξ_{nk} , $k \leq k(n)$ — последовательность наборов случайных величин, удовлетворяющая следующим условиям:

- 1) случайные величины ξ_{nk} независимы между собой для каждого n ;
- 2) $\mathbb{E}\xi_{nk} = 0$ для всех n и k ;
- 3) $\sum_{k=1}^{k(n)} \mathbb{E}\xi_{nk}^2 = 1$ для каждого n ;
- 4) $\lim_n \sum_{k=1}^{k(n)} \mathbb{E}\xi_{nk}^2 I_{\{|\xi_{nk}| > C\}} = 0$ для всех $C > 0$.

Тогда распределения случайных величин $\sum_{k=1}^{k(n)} \xi_{nk}$ слабо сходятся к нормальному распределению с параметрами 0 и 1, т.е. $N(0, 1)$.

Доказательство. Обозначим

$$\varphi_{nk}(t) := \mathbb{E}e^{it\xi_{nk}} = \varphi_{\xi_{nk}}(t)$$

Ввиду независимости ξ_{nk} нам достаточно доказать сходимость

$$\lim_n \prod_{k=1}^{k(n)} \varphi_{nk}(t) = e^{-t^2/2}$$

или, что то же

$$\lim_n \sum_{k=1}^{k(n)} \ln(\varphi_{nk}(t)) = -t^2/2. \quad (1)$$

Доказательство разбивается на ряд этапов. Мы стремимся избавиться в (1) от \ln и заменить $\ln(\varphi_{nk}(t))$ на $\varphi_{nk}(t) - 1$. Для этого мы воспользуемся неравенством с), где $x = \varphi_{nk}(t) - 1$. Чтобы иметь право это сделать, нам надо доказать, что начиная с некоторого n все $\varphi_{nk}(t) - 1$ будут меньше $1/2$. Для этого мы докажем, что

$$\max_k |\varphi_{nk}(t) - 1| \rightarrow 0. \quad (2)$$

Лемма 1.

$$\lim_n \max_k \mathbb{E} \xi_{nk}^2 = 0. \quad (3)$$

Доказательство.

$$\begin{aligned} \max_k \mathbb{E} \xi_{nk}^2 &\leq \max_k \mathbb{E} \xi_{nk}^2 I_{\{|\xi_{nk}| \leq C\}} + \mathbb{E} \xi_{nk}^2 I_{\{|\xi_{nk}| > C\}} \leq \\ &\leq \max_k \mathbb{E} \xi_{nk}^2 I_{\{|\xi_{nk}| \leq C\}} + \sum_k^{k(n)} \mathbb{E} \xi_{nk}^2 I_{\{|\xi_{nk}| > C\}}. \end{aligned}$$

Первое слагаемое оценивается числом C^2 , которое может быть выбрано сколь угодно малым, а второе слагаемое, согласно условию с), сходится к 0. Это доказывает (3).

Теперь мы докажем (2):

Лемма 2.

$$\max_k |\varphi_{nk}(t) - 1| = \max_k |\mathbb{E} e^{it\xi_{nk}} - 1| =$$

(согласно 2) мы можем представить 0 как $\mathbb{E} \xi_{nk}$)

$$= \max_k |\mathbb{E} (e^{it\xi_{nk}} - it\xi_{nk} - 1)| \leq$$

(согласно б))

$$\leq \max_k |\mathbb{E} \xi_{nk}^2| \rightarrow 0$$

согласно Лемме 1. Итак, мы заменяем $\sum_{k=1}^{k(n)} \ln(\varphi_{nk}(t))$ на $\sum_{k=1}^{k(n)} (\varphi_{nk}(t) - 1)$, но нам надо доказать, что ошибка, которая происходит от этой замены, сходится к 0:

$$\sum_{k=1}^{k(n)} |\ln(\varphi_{nk}(t)) - (\varphi_{nk}(t) - 1)| \leq \sum_{k=1}^{k(n)} |\varphi_{nk}(t) - 1|^2$$

(квадрат мы представим как произведение и оценим первый множитель максимумом)

$$\leq \max_k |\varphi_{nk}(t) - 1| \sum_{k=1}^{k(n)} |\varphi_{nk}(t) - 1|.$$

Согласно лемме 2, первый множитель сходится к 0, а второй множитель оценивается

$$\sum_{k=1}^{k(n)} |\varphi_{nk}(t) - 1| = \sum_{k=1}^{k(n)} |\mathbb{E} (e^{it\xi_{nk}} - it\xi_{nk} - 1)| \leq \sum_{k=1}^{k(n)} \mathbb{E} \xi_{nk}^2 = 1.$$

Итак, нам осталось доказать:

$$\lim_n \sum_{k=1}^{k(n)} (\varphi_{nk}(t) - 1) = -t^2/2. \quad (4)$$

Опять мы используем представление 0 в 2):

$$\sum_{k=1}^{k(n)} (\varphi_{nk}(t) - 1) = \sum_{k=1}^{k(n)} \mathbb{E} (e^{it\xi_{nk}} - it\xi_{nk} - 1) =$$

(далее мы используем формулу замены переменной)

$$\begin{aligned}
&= \int_{\mathbf{R}} (e^{itx} - itx - 1) d \sum_{k=1}^{k(n)} \mathbb{P}_{\xi_{nk}}(x) = \\
&= \int_{\mathbf{R}} \frac{e^{itx} - itx - 1}{x^2} x^2 d \sum_{k=1}^{k(n)} \mathbb{P}_{\xi_{nk}}(x) = \int_{\mathbf{R}} \frac{e^{itx} - itx - 1}{x^2} d \sum_{k=1}^{k(n)} \mathbb{E} \xi_{nk}^2 I_{\{\xi_{nk} < x\}}.
\end{aligned}$$

Над этими равенствами надо подумать. Функция $f(x) = \frac{e^{itx} - itx - 1}{x^2}$ сходится к 0 на $\pm\infty$ и может иметь особенность лишь в 0. Но эта особенность легко раскрывается и предел $\lim_{x \rightarrow 0} = -t^2/2$. Второе равенство мы запишем для одного слагаемого:

$$\int_{\mathbf{R}} g(x) x^2 d\mathbb{P}_{\xi} = \int_{\mathbf{R}} g(x) d\mathbb{E} \xi^2 I_{\{\xi < x\}}, \quad (5)$$

оно верно, если g непрерывна и ограничена, а $\mathbb{E} \xi^2 < \infty$. (Оба интеграла можно со сколь угодно малой погрешностью заменить на \int_a^b , а эти интегралы приблизить суммами Римана.)

Итак, равенство (5) достаточно доказать лишь для $g(x) = I_{[a,b)}(x)$. Но для этого случая оно очевидно по формуле замены переменной:

$$\int_{\mathbf{R}} I_{[a,b)}(x) x^2 d\mathbb{P}_{\xi}(x) = \mathbb{E} \xi^2 I_{\{a \leq \xi < b\}} = \int_{\mathbf{R}} I_{[a,b)}(x) d\mathbb{E} \xi^2 I_{\{\xi < x\}}.$$

Мы имеем последовательность функций распределения

$$F_n(x) = \sum_{k=1}^{n(k)} \mathbb{E} \xi_{nk}^2 I_{\{\xi_{nk} < x\}}.$$

Действительно, $\lim_{x \rightarrow \infty} F_n(x) = 1$ согласно 3), $\lim_{x \rightarrow -\infty} F_n(x) = \mathbb{E} \xi_{nk}^2 I_{\emptyset} = 0$. Осталось доказать, что F_n слабо сходится к функции распределения вероятности $\delta_{\{0\}}$, сосредоточенной в точке 0. Эта функция распределения равна 0 для $x < 0$ и 1 для $x > 0$. Имеем:

$$\lim_n \sum_{k=1}^{k(n)} \mathbb{E} \xi_{nk}^2 I_{\{\xi_{nk} < x\}} \leq \lim_n \sum_{k=1}^{k(n)} \mathbb{E} \xi_{nk}^2 I_{\{|\xi_{nk}| > x\}} = 0$$

для $x < 0$, и

$$\lim_n \sum_{k=1}^{k(n)} \mathbb{E} \xi_{nk}^2 I_{\{\xi_{nk} < x\}} = 1 - \lim_n \sum_{k=1}^{k(n)} \mathbb{E} \xi_{nk}^2 I_{\{\xi_{nk} > x\}} = 1$$

для $x > 0$. Резюмируем:

$$\int_{\mathbf{R}} f(x) d \sum_{k=1}^{n(k)} \mathbb{E} \xi_{nk}^2 I_{\{\xi_{nk} < x\}} \rightarrow \int_{\mathbf{R}} f(x) d\delta_{\{0\}} = f(0) = -t^2/2.$$

Следствие. Цpt для одинаково распределенных случайных величин является частным случаем теоремы Линдеберга.

Для этого нам нужно рассмотреть в условиях Центральной предельной теоремы из §13 последовательность случайных серий

$$\xi_{nk} = \frac{\xi_k - m}{\sigma \sqrt{n}}, \quad k \leq n.$$

Выполнение условий 1)-3) очевидно. Надо проверить выполнение условия 4).

$$\begin{aligned} & \lim_n \sum_{k=1}^n \mathbb{E} \frac{(\xi_k - m)^2}{\sigma^2 n} I_{\left\{\left|\frac{\xi_k - m}{\sigma \sqrt{n}}\right| > C\right\}} = \\ & = \lim_n \mathbb{E} \frac{(\xi_1 - m)^2}{\sigma^2} I_{\left\{\left|\frac{\xi_1 - m}{\sigma}\right| > C\sqrt{n}\right\}} \rightarrow 0. \end{aligned}$$

Здесь мы использовали одинаковость распределений и равенство слагаемых в сумме, а также

$$\left\{\left|\frac{\xi_1 - m}{\sigma}\right| > C\sqrt{n}\right\} \rightarrow \emptyset (n \rightarrow \infty).$$

§15. Применения предельных теорем. Многомерные предельные теоремы

Нормальную аппроксимацию можно использовать для приближенного вычисления вероятности $\mathbb{P}\{S_n < k\}$, где S_n — число гербов в n подбрасываниях монеты с параметром p , близким к середине интервала $(0, 1)$. Если параметр p близок к 0, а k небольшое, то можно использовать для приближений распределение Пуассона. Нормальное приближение можно использовать также для распределений сумм различных независимых случайных величин (теорема Линденберга). Все это должно быть усвоено на практике. Нормальное приближение возможно также для распределения χ_n^2 , причем для некоторых значений n можно проверить эффективность нормального приближения, сравнив с табличными значениями. Для распределения Стьюдента также можно использовать нормальное приближение, но обосновывается это законом больших чисел, согласно которому знаменатель в представлении t_n сходится к 1.

Если S — сумма большого числа случайных величин примерно одного порядка со средним и дисперсией, то мы нормируем и центрируем эту случайную величину и используем таблицы функции распределения (обычно обозначается Φ) стандартной нормальной случайной величины. А именно,

$$\mathbb{P}\{S < a\} = \mathbb{P}\left\{\frac{S - \mathbb{E}(S)}{\sqrt{\mathbb{D}(S)}} < \frac{a - \mathbb{E}(S)}{\sqrt{\mathbb{D}(S)}}\right\} = \Phi\left(\frac{a - \mathbb{E}(S)}{\sqrt{\mathbb{D}(S)}}\right).$$

Аналогично поступаем, если нас интересуют вероятности других событий, связанных с S . Таблицы функции Φ как правило приводятся лишь для положительных x . Для вычисления значения функции Φ на отрицательной полуоси используйте четность функции плотности $N(0, 1)$.

1. Моделирование нормального распределения.

Мы будем обозначать через r или r_n псевдослучайные числа, моделируемые компьютером с помощью программного пакета, предполагается, что они независимы для разных i и имеют равномерное распределение на отрезке $[0, 1]$.

В принципе, распределение случайной величины с любой функцией распределения F может моделироваться случайной величиной $F^{-1}(r)$. Проверим:

$$\mathbb{P}\{F^{-1}(r) < x\} = \mathbb{P}\{r < F(x)\}.$$

Но этот универсальный метод не очень хорош для нормального распределения, так как обратная функция к функции Φ не является элементарной, ее вычисление достаточно трудоемко.

Одним из способов моделирования нормального распределения является принятие в качестве $N(0, 1)$ -нормальной случайной величины ξ суммы $r_1 + \dots + r_{12} - 6$, где r_i — случайные числа, которые предполагаются равномерно распределенными на отрезке $[0, 1]$. По центральной предельной теореме ξ примерно нормально, мы знаем, что среднее ξ равно 0, а дисперсия — 1.

Другой способ позволяет одновременно моделировать пару $N(0, 1)$ -нормальных случайных величин ξ_1 и ξ_2 . Точнее, мы моделируем координаты (ρ, φ) вектора (ξ_1, ξ_2) в полярной системе

координат. Оказывается, что эти координаты независимы и обратная к функции распределения каждой из них легко вычисляется. Действительно, перейдем к полярным координатам в интеграле

$$\mathbb{P}\{(\xi_1, \xi_2) \in B\} = \int \int_B \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dx dy = \int \int_B \frac{1}{2\pi} e^{-\frac{\rho^2}{2}} \rho d\rho d\varphi.$$

Мы видим, что подынтегральная функция распадается в произведение функций от ρ и от φ , φ имеет равномерное распределение на отрезке $[0, 2\pi]$ и поэтому легко моделируется, а ρ имеет функцию распределения

$$\mathbb{P}\{\rho < x\} = \int_0^x e^{-\frac{\rho^2}{2}} \rho d\rho = 1 - e^{-x^2/2}.$$

Обратная функция выражается через элементарные функции.

Третий наиболее practicalnyj способ предложил Knuth. Он рассмотрел график плотности $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ и покрыл почти всю площадь под этим графиком (она равна 1) несколькими прямоугольниками. В результате почти вся вероятность (за исключением очень маленького ε) моделируется просто, каждому прямоугольнику соответствует отрезок в $[0, 1]$, длина которого равна площади прямоугольника, и в случае попадания r в этот отрезок мы выбираем соответствующую r точку на проекции прямоугольника на ось x -ов. Если не повезло, и компьютер выбрал r , соответствующий остаточной вероятности величины ε , то компьютер проводит очень трудоемкие вычисления обратных функций.

2. Многомерные характеристические функции.

Определение. Характеристическая функция φ_{ξ} случайного вектора $\bar{\xi} = (\xi_1, \dots, \xi_n)$ или, что то же, совместная характеристическая функция $\varphi_{\xi_1, \dots, \xi_n}$ задается соотношением

$$\varphi_{\xi_1, \dots, \xi_n}(t_1, \dots, t_n) = \mathbb{E}e^{i(t_1\xi_1 + \dots + t_n\xi_n)}.$$

У нее те же свойства, что и у одномерной характеристической функции, в том числе теоремы единственности и непрерывности, но доказываются они сложнее, мы этого делать не будем. Что касается используемой в доказательстве теоремы единственности многомерной формулы обращения, то в ней речь идет об интегральном представлении меры

$$\mathbb{P}\left\{\prod_{i=1}^n [a_i, b_i]\right\},$$

причем аналогом условия непрерывности функции распределения в точке служит свойство

$$\mathbb{P}\left\{\bigcup_{i=1}^n \{(x_1, \dots, x_n) : x_i = a_i\} \bigcup \bigcup_{i=1}^n \{(x_1, \dots, x_n) : x_i = b_i\}\right\} = 0.$$

Как и в одномерном случае, показывается, что множество пар точек $\{(a_i), (b_i)\}$, для которых это свойство нарушается, не более чем счетно.

Упражнение. Покажите, что из независимости случайных величин ξ_1, \dots, ξ_n следует

$$\varphi_{\xi_1, \dots, \xi_n}(t_1, \dots, t_n) = \varphi_{\xi_1}(t_1) \cdot \dots \cdot \varphi_{\xi_n}(t_n). \quad (6)$$

используя многомерную теорему единственности, покажите, что обратно, из равенства (*) следует независимость ξ_1, \dots, ξ_n .

Как и в одномерном случае, слабая сходимость многомерных распределений означает сходимость интегралов от всех непрерывных ограниченных функций, заданных на пространстве той же размерности.

Как и в одномерном случае, мы говорим о сходимости по распределению $\bar{\xi}_n$ к $\bar{\xi}$ и будем обозначать

$$\bar{\xi}_n \xrightarrow{\mathcal{D}} \bar{\xi},$$

если

$$\mathbb{P}_{\bar{\xi}_n} \xrightarrow{\text{сл}} \mathbb{P}_{\bar{\xi}},$$

Замечание. Из упражнения следует, что если

$$\bar{\xi}_n \xrightarrow{\mathcal{D}} \bar{\xi},$$

$\bar{\xi}_n = (\xi_n^{(1)}, \dots, \xi_n^{(d)})$, $\bar{\xi} = (\xi^{(1)}, \dots, \xi^{(d)})$, случайные величины $\xi_n^{(1)}, \dots, \xi_n^{(d)}$ независимы для любого n , то случайные величины $\xi^{(1)}, \dots, \xi^{(d)}$ независимы. Это тем более верно, если

$$\bar{\xi}_n \xrightarrow{\mathbb{P}} \bar{\xi}.$$

Сходимость случайных векторов по вероятности

$$\bar{\xi}_n \xrightarrow{\mathbb{P}} \bar{\xi}$$

означает

$$\mathbb{P}\{\|\bar{\xi}_n - \bar{\xi}\| > \varepsilon\} \rightarrow 0$$

для любого $\varepsilon > 0$, где $\|\cdot\|$ — норма в \mathbf{R}^d .

Замечание. Если $\bar{\xi}_n, \bar{\xi}$ — последовательность r -мерных случайных векторов, f — непрерывная функция на \mathbf{R}^r , и имеет место сходимость

$$\mathbb{P}_{\bar{\xi}_n} \xrightarrow{\text{сл}} \mathbb{P}_{\bar{\xi}},$$

то

$$\mathbb{P}_{f(\bar{\xi}_n)} \xrightarrow{\text{сл}} \mathbb{P}_{f(\bar{\xi})}.$$

Доказательство немедленно вытекает из определений.

Посмотрим, как меняется совместная характеристическая при линейном преобразовании случайного вектора (ξ_1, \dots, ξ_n) — переходе от вектора (ξ_1, \dots, ξ_n) к вектору $(\xi_1, \dots, \xi_n)A$, где $A = [a_{ij}]$ задается матрицей. Мы рассматриваем вместе с вектором (ξ_1, \dots, ξ_n) также (t_1, \dots, t_n) и представляем $t_1\xi_1 + \dots + t_n\xi_n$ как произведение вектор-строки на вектор-столбец

$$t_1\xi_1 + \dots + t_n\xi_n = (\xi_1, \dots, \xi_n)(t_1, \dots, t_n)'.$$

Если мы переходим к случайному вектору $(\xi_1, \dots, \xi_n)A$, то получаем

$$(\xi_1, \dots, \xi_n)A(t_1, \dots, t_n)' = (\xi_1, \dots, \xi_n)((t_1, \dots, t_n)A')'.$$

Итак,

$$\varphi_{(\xi_1, \dots, \xi_n)A}(t_1, \dots, t_n) = \varphi_{(\xi_1, \dots, \xi_n)}((t_1, \dots, t_n)A').$$

Применим это преобразование к стандартному нормальному случайному вектору (ξ_1, \dots, ξ_n) . Характеристическая функция у этого вектора равна

$$\varphi_{(\xi_1, \dots, \xi_n)}(t_1, \dots, t_n) = e^{-\frac{1}{2}(t_1, \dots, t_n)(t_1, \dots, t_n)'}$$

Вспомним правило транспонированных матрицы произведения, получаем

$$\varphi_{(\xi_1, \dots, \xi_n)A}(t_1, \dots, t_n) = e^{-\frac{1}{2}(t_1, \dots, t_n)A'(t_1, \dots, t_n)'} = .$$

$$e^{-\frac{1}{2} \sum_{k,l} \mu_{kl} t_k t_l},$$

где $\mu_{kl} = \sum_i a_{ik} a_{il}$, поэтому матрица $[\mu_{kl}]$ неотрицательно определена как произведение матрицы и ее транспонированной.

В нецентрированном случае мы имеем характеристическую функцию

$$\varphi_{\eta_1, \dots, \eta_n}(t_1, \dots, t_n) = e^{\sum_k i m_k t_k - \frac{1}{2} \sum_{k,l} \mu_{kl} t_k t_l}. \quad (7)$$

Эта характеристическая функция дает самое общее определение многомерного нормального распределения. Например, если матрица $[\mu_{kl}]$ является вырожденной, то мы имеем вырожденное нормальное распределение, которое сосредоточено на некотором линейном многообразии, а следовательно, не имеет плотности.

Как и в одномерном случае, частные производные характеристической функции в нуле определяют средние $\mathbb{E}(\eta_k)$ случайных величин η_k , умноженные на i , а вторые производные равны смешанным моментам с отрицательным знаком:

$$-\mathbb{E}\{(\eta_k - \mathbb{E}(\eta_k))(\eta_l - \mathbb{E}(\eta_l))\} = \mu_{kl}.$$

Упражнение 20. а. Используя предыдущее замечание, покажите, что в (7)

$$m_k = \mathbb{E}(\eta_k), \mu_{kl} = \mathbb{E}\{(\eta_k - \mathbb{E}(\eta_k))(\eta_l - \mathbb{E}(\eta_l))\}.$$

б. Дано разложение $A_1 + \dots + A_r = \Omega$, где $\mathbb{P}(A_k) = p_k$. Случайные величины ξ_k задаются соотношением $\xi_k(\omega) = I_{A_k}$. Вычислите совместную характеристическую функцию вектора (ξ_1, \dots, ξ_r) , а также матрицы смешанных моментов и центральных смешанных моментов.

3. Многомерная центральная предельная теорема.

В многомерном случае не всегда можно нормировать случайные векторы некоторой матрицей, поэтому формулировка центральной предельной теоремы выглядит следующим образом:

Теорема. Пусть (ξ_n) – последовательность независимых одинаково распределенных случайных векторов с нулевыми средними и матрицей вторых моментов $[\mu_{kl}]$. Тогда распределение случайного вектора

$$\frac{\bar{\xi}_1 + \dots + \bar{\xi}_n}{\sqrt{n}}$$

слабо сходится к нормальному распределению с нулевыми средними и той же матрицей вторых моментов $[\mu_{kl}]$.

Эта теорема работает и в случае сходимости к вырожденному нормальному распределению, и именно в этой ситуации мы ее будем использовать. От средних надо избавляться в этой теореме, так как при суммировании случайных величин средние складываются, величина суммы пропорциональна n , а делим мы сумму лишь на \sqrt{n} , в результате нормированное число \sqrt{n} среднее стремится к бесконечности.

Доказательство теоремы мы опускаем.

Задача. Найти слабый предел последовательности распределений $\frac{\xi_1 + \dots + \xi_n}{\eta_1 + \dots + \eta_n}$, где ξ_i, η_j независимы и имеют одинаковое распределение со средним 0 и дисперсией 1.

Часть 4. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Введение

Задача математической статистики — по результатам нескольких экспериментов (или наблюдений) максимально точно восстановить некоторые из условий проведения эксперимента. Ниже я приведу определение из авторитетного источника, но заучивать это определение не стоит.

Математическая статистика — раздел математики, посвященный математическим методам сбора, систематизации, обработки и интерпретации статистических данных, а также использованию их для научных или практических выводов. Правила и процедуры математической статистики опираются на *теорию вероятностей*, позволяющую оценить точность и надежность выводов, получаемых в каждой задаче на основе имеющегося статистического материала.

Из статьи *Математическая статистика. И.А. Ибрагимов, Ю.В. Прохоров, Н.Н. Ченцов, Д.М. Чубисов* в энциклопедии "Вероятность и математическая статистика" (Изд-во БРЭ, 1999).

При работе с таблицами и доведения решения до числа часто используется термин *p-квантиль распределения*. Так называется функция, обратная к функции распределения. Если F — функция распределения, то *p*-квантиль — это такое число x , что $F(x) = p$. Причем *p* задаем мы сами (заказчик работы или преподаватель), а x используется для выработки ответа.

§16. Выборка.

1. Пример. Пусть имеется партия деталей из m деталей, содержащая m_1 деталей 1-го сорта, m_2 деталей 2-го сорта, m_3 деталей 3-го сорта ($m = m_1 + m_2 + m_3$). Эксперимент состоит в выборе (с возвращением) из m деталей n деталей и проверке их качества. В i -м эксперименте наблюдается число x_i — сорт i -й детали. Задача состоит в максимально точной оценке чисел m_1, m_2, m_3 . Очевидно, что для этой оценки мы имеем в нашем распоряжении лишь числа x_1, x_2, \dots, x_n , набор этих чисел называется выборкой из генеральной совокупности, генеральная совокупность в этом примере — партия из m деталей. Очевидно, что хорошей оценкой чисел m_1, m_2, m_3 являются числа $n_1(m/n), n_2(m/n), n_3(m/n)$, столь же очевидно, что эти числа как правило не равны числам m_1, m_2, m_3 (более того, эти числа могут быть дробными). Поэтому возникает вопрос о величине ошибки при нашей оценке, о вероятности ошибки.

2. Следуя статистической традиции, мы сохраним название *выборка* и в том случае, когда результаты экспериментов x_1, x_2, \dots, x_n получены по-другому, при этом под генеральной совокупностью будет пониматься множество всех возможных значений эксперимента с заданной на нем некоторой вероятностной мерой, вероятность нахождения результата эксперимента в некотором подмножестве генеральной совокупности равна вероятности данного подмножества. Описание указанной вероятности (или некоторых ее свойств) и является задачей математической статистики. Генеральная совокупность в такой обобщенной трактовке — не обязательно конечно множество (она может быть равна \mathbf{N} , \mathbf{R} , \mathbf{R}^2 и т.д. с какими-то неизвестными нам вероятностями на подмножествах. Выборка принимает свои значения в генеральной совокупности). Почти всегда мы будем считать, что генеральная совокупность — множество вещественных чисел, на котором задана некоторая борелевская вероятность.

3. Выборочные характеристики.

С выборкой $\{x_1, x_2, \dots, x_n\}$ принято связывать ряд характеристик, которые называются эмпирическими или выборочными:

Эмпирическое распределение — вероятность на \mathbf{R} , которая получится, если каждому элементу выборки мы припишем одну и ту же вероятность, равную $1/n$ (единица, деленная на число элементов выборки), обозначаем эмпирическое распределение — $\mathbb{P}_{\mathcal{E}}$.

Эмпирическая функция распределения — функция распределения эмпирического распределения, т.е.

$F_{\Theta}(x) = \frac{\{\text{число } x_i, \text{ меньших данного } x\}}{n}$ для всех x .

Эмпирическое среднее:

$$\bar{x} = \int x dF_{\Theta}(x) = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

Эмпирическая дисперсия:

$$s^2(\text{ или } s_x^2) := \int x^2 dF_{\Theta}(x) - \left(\int x dF_{\Theta}(x) \right)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (\sigma)$$

Обратите внимание на последнее равенство в (σ) , которое мы будем использовать как тождество (следствие известного тождества

$$\mathbb{D}\xi = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2 = \mathbb{E}(\xi - \mathbb{E}\xi)^2$$

применительно к эмпирической вероятности).

Эмпирические характеристики имеют смысл и тогда, когда генеральная совокупность не одномерна. Пусть, например, выборка — набор векторов $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, в этом случае эмпирическое распределение задано на \mathbf{R}^2 формулой $\mathbb{P}_{\Theta}\{(x_i, y_i)\} = 1/n$ для всех $i \leq n$, точно так же определяются $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$, \bar{y} , s_x^2 и s_y^2 , в этом случае вводится эмпирический коэффициент корреляции

$$r_{x,y} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(s_x^2)^{1/2}(s_y^2)^{1/2}}$$

и т.д.

Упражнение. Запишите формулу для вычисления по выборке $\{x_1, x_2, \dots, x_n\}$ эмпирического третьего момента и эмпирического четвертого центрального момента.

Упражнение. Придумайте конкретную выборку, состоящую из четырех чисел, постройте по ней эмпирическую функцию распределения, вычислите \bar{x} и s_x^2 .

§17. Вероятностная и статистическая модели

Обозначим через $\mathbb{P}_{\text{ист}}$ истинное распределение вероятностей на данной одномерной генеральной совокупности, наша задача — как-то, используя результаты экспериментов x_1, x_2, \dots, x_n , приблизить $F_{\text{ист}}$. Стандартный прием (если мы знаем, что результатом эксперимента может быть любое вещественное число) нарисовать эмпирическую функцию распределения F_{Θ} и сладить ее, используя лекала. Очень давно статистики-практики так и поступали. Недостатки этого приема (методом его назвать трудно) очевидны — необоснованность выводов, необходимость большого числа наблюдений (иначе эмпирическая функция распределения будет иметь слишком большие разрывы). Бывают случаи, когда в распоряжении статистика по-существу нет иного приема. Ситуация существенно улучшается, если статистик имеет некую априорную информацию о виде распределения $\mathbb{P}_{\text{ист}}$ (например, что $\mathbb{P}_{\text{ист}}$ — нормальное распределение, или экспоненциальное, или равномерное, или является распределением Бернуlli, например, в примере в пункте 1 из условий задачи видно, что распределение $\mathbb{P}_{\text{ист}}$ сосредоточено в точках $\{1, 2, 3\}$). В такой ситуации задача существенно облегчается, мы можем считать, что $\mathbb{P}_{\text{ист}}$ принадлежит некоторому классу распределений \mathcal{P} , который называется *вероятностной моделью* решаемой статистической задачи.

1. Примеры вероятностных моделей.

Нормальная модель $\mathcal{P} = \{P_{m,\sigma}\}$, где $P_{m,\sigma}$ — нормальное распределение с параметрами m и σ , эта модель задается набором плотностей распределений

$$p_{m,\sigma}(x) = \frac{1}{\sigma(2\pi)^{1/2}} e^{-\frac{1}{2}(\frac{x-m}{\sigma})^2}, x \in \mathbf{R}, \text{ где } m \in \mathbf{R}, \sigma \geq 0.$$

Экспоненциальная модель

$$p_\lambda(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Равномерная модель

$$p_{m,\Delta}(x) = \begin{cases} 1/(2\Delta), & \text{если } x \in (m - \Delta, m + \Delta); \\ 0 & \text{в противном случае.} \end{cases}$$

Модель Коши

$$p_{a,m}(x) = \frac{a}{\pi(a^2 + (x - m)^2)}$$

Модель Пуассона

$$\mathbb{P}_\lambda\{k\} = e^{-\lambda} \frac{\lambda^k}{k!}, k = 0, 1, 2, \dots$$

Модель Бернуlli

$$\mathbb{P}_p\{1\} = p, P_p\{0\} = 1 - p.$$

Другие модели (некоторые из них представляют большой практический интерес) в курсе не рассматривается.

В рамках используемой нами вероятностной модели результаты наблюдений x_1, x_2, \dots, x_n интерпретируются как наблюдения n независимых одинаково распределенных случайных величин, причем их одинаковое распределение есть $\mathbb{P} = \mathbb{P}_{\text{ист}}$ и оно нам неизвестно, но известно, что оно принадлежит данному классу \mathcal{P} . При этих предположениях по данному \mathbb{P} можно построить совместное распределение $\mathbb{P}^{(n)}$ случайных величин x_1, x_2, \dots, x_n , заданное на пространстве \mathbf{R}^n . Напомним, что если x_1, x_2, \dots, x_n — случайные величины, имеющие одинаковую плотность p , то их совместная функция плотности равна

$$p^{(n)}(x_1, x_2, \dots, x_n) = p(x_1)p(x_2)\dots p(x_n), x_i \in \mathbf{R}.$$

Итак, по вероятностной модели \mathcal{P} строится класс распределений на n -мерном пространстве $\mathcal{P}^{(n)} = \{P^{(n)} : P \in \mathcal{P}\}$, который называется *статистической моделью*.

2. Примеры статистических моделей.

Нормальная статистическая модель задается классом плотностей

$$p_{m,\sigma}^{(n)}(x_1, x_2, \dots, x_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2}\sum_{i=1}^n \left(\frac{x_i - m}{\sigma}\right)^2\right], x_i \in \mathbf{R}$$

где m пробегает \mathbf{R} , а $\sigma \in (0, \infty)$.

Пуассоновская статистическая модель задается классом распределений

$$\mathbb{P}_\lambda^{(n)}\{k_1, k_2, \dots, k_n\} = \exp(-n\lambda) \frac{\lambda^{k_1+k_2+\dots+k_n}}{k_1!k_2!\dots k_n!}, k_i \in \{0\} \cup \mathbf{N},$$

где λ пробегает $(0, \infty)$.

Модель Бернуlli задается классом распределений

$$\mathbb{P}_p^{(n)}\{\delta_1, \delta_2, \dots, \delta_n\} = p^{\sum_{i=1}^n \delta_i} (1-p)^{\sum_{i=1}^n (1-\delta_i)}, \delta_i \in \{0, 1\},$$

где p пробегает все числа в $(0, 1)$.

Упражнение. Постройте равномерную, экспоненциальную, Коши статистические модели.

Использование вероятностной и статистической модели позволяет придать точный смысл понятию 'вероятность ошибки', оценивать качества различных статистических критериев. В

далнейшем делать статистические выводы мы будем в рамках той или иной вероятностно-статистической модели. При этом, конечно же, возникает опасность ошибиться в выборе модели (что приведет к неправильным статистическим выводам). В простейшей ситуации, когда результаты наблюдений могут принимать лишь два значения — 0 или 1, такой опасности нет, однозначно выбирается модель Бернулли, в более сложных ситуациях правильность выбора модели зависит от искусства статистика, его опыта, кроме того, математическая статистика предлагает методы проверки правильности выбора модели.

3. Параметрические модели. Все рассмотренные выше модели входят в класс параметрических моделей, параметрическая модель имеет вид $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, где θ пробегает некоторое множество значений параметра Θ (не обязательно \mathbf{R}), если множество Θ двумерно или трехмерно, то модель называется двухпараметрической или трехпараметрической.

Упражнение. Указать, какие из построенных моделей являются однопараметрическими, и какие — двухпараметрическими.

Замечание. В более сложных задачах могут использоваться и более сложные вероятностные модели. Например, если генеральная совокупность является двумерной, то и вероятностная модель состоит из распределений вероятностей на \mathbf{R}^2 , в этом случае статистическая модель состоит из распределений на \mathbf{R}^{2n} .

4. Статистики.

Статистикой называется любое измеримое отображение T из пространства выборок в \mathbf{R} . Примеры статистик:

$$T(x_1, x_2, \dots, x_n) = x_i; T(x_1, x_2, \dots, x_n) = \bar{x}, s^2, \max(x_1, x_2, \dots, x_n)$$

и т.д.

Обозначение: $\mathbb{E}_\theta T = \int T(x_1, x_2, \dots, x_n) dP_\theta^{(n)}(x_1, x_2, \dots, x_n)$.

Так как x_1, x_2, \dots, x_n рассматриваются нами в рамках вероятностной модели как независимые одинаково распределенные случайные величины, распределение которых неизвестно, но принадлежит классу \mathcal{P} , то $T(x_1, x_2, \dots, x_n)$ как функция случайных величин, также является случайной величиной, распределение которой принадлежит некоторому классу, который строится по классу \mathcal{P} . В случае параметрической модели $\{P_\theta\}_{\theta \in \Theta}$ используется обозначение $P_{\theta, T}$ — распределение статистики T при условии, что значение параметра равно θ . Одной из предварительно решаемых задач при статистическом выводе является вывод распределений используемых нами статистик (разумеется, для всех возможных значений параметра).

Упражнение. а) Покажите, что распределение статистики

$$T(x_1, x_2, \dots, x_n) = (x_1 - x_2)/(x_3 - x_4)$$

не зависит от значений параметров m и σ в нормальной модели. б) Найдите распределение этой статистики.

5. Основные задачи математической статистики.

- 1) Задача оценки параметра в рамках параметрической вероятностной модели.
- 2) Задача сравнения двух гипотез (т.е. каждая из гипотез означает свою вероятностную модель \mathcal{P}_1 или \mathcal{P}_2 , нужно сделать выбор между этими моделями).
- 3) Задача проверки гипотезы (т.е., дана вероятностная модель \mathcal{P} , мы должны проверить, согласуются ли результаты эксперимента с этой моделью).

Возможны и другие задачи.

Давайте еще раз сформулируем, что такое — вероятностная модель, и что такое — статистическая модель. Это будет не математическое, а словесное определение.

Итак, вероятностная модель — это класс вероятностных распределений на пространстве возможных результатов наблюдений. Статистическая модель — это класс распределений на пространстве выборок, который строится по вероятностной модели в предположении независимости и одинаковости экспериментов. Как мы уже видели, пространство в статистической

модели — это n -я степень пространства в вероятностной модели, а вероятность в статистической модели получается из соответствующей вероятности в вероятностной модели по правилу произведения одинаковых функций, но взятых в разных точках области определения.

Не всегда вероятностная модель — \mathbf{R} , а статистическая — \mathbf{R}^n . Выборки могут быть и многомерными, например, в двумерном случае пространство выборок имеет вид $(\mathbf{R}^2)^n$. В статистике случайных процессов, если траектории непрерывны, то результаты наблюдений — элементы пространства непрерывных функций с известной вам равномерной нормой супремума. А если траектории процесса разрывны, то для пространства результатов наблюдений используется метрика Скорохода, в которой функции с одинаковыми разрывами в близкие моменты времени считаются близкими.

Наконец, пространство результатов наблюдений состоит из двух точек (как правило, это $\{0, 1\}$) в случае наблюдения в каждом эксперименте одного и того же события A . Мы считаем результат эксперимента равным 1, если событие A произошло, и равным 0, если оно не произошло.

Заметим, что решение задачи 2) сравнения гипотез используется для задачи проверки гипотезы в ситуации, когда этой гипотезе сопоставляется некая альтернатива.

Для решения этих задач используются специально подобранные статистики. Они сами являются случайными величинами и имеют распределение в условиях фиксированной вероятностной модели. Важнейшей общей задачей математической статистики является нахождение распределения статистики. В условиях параметрической модели для задачи проверки гипотезы особый интерес представляют статистики, распределение которых не зависит от неизвестного нам значения параметра θ . В задаче оценивания параметра статистика должна представлять информацию о параметре, поэтому ее распределение наоборот должно зависеть от выбора параметра модели, однако желательно наличие функции, $f(T, \theta)$, используемой статистики и параметра, распределение которой не зависит от параметра. Обо всем этом мы поговорим в дальнейшем.

§18. Оценка параметров

1. Оценкой параметра θ в параметрической модели $\{P_\theta\}_{\theta \in \Theta}$ называется измеримое отображение пространстве выборок в множество параметров. В простейшей ситуации, когда наблюдения являются одномерными, а множество параметров содержится в \mathbf{R} , это — некоторая статистика

$$\hat{\theta} : \mathbf{R}^n \rightarrow \Theta.$$

Оценкой можно назвать любую, в том числе и самую неразумную функцию (например, $\hat{\theta}(x_1, x_2, \dots, x_n) = \theta_0$, т.е. не зависит от результатов наблюдений). Ниже мы введем свойства оценок, наличие которых делают оценки разумными.

2. Несмешенная оценка. Мы приведем два определения несмешенной оценки, если вдуматься в смысл этих определений, это одно и то же.

Определение 1. Оценка $\hat{\theta}$ называется *несмешенной*, если ее среднее значение совпадает с истинным значением параметра $\mathbb{E}\hat{\theta} = \theta_{\text{ист}}$.

Замечание 1. Напомним (см. лекцию 1, п. 6), что $\hat{\theta}$ — случайная величина, функция независимых одинаково распределенных случайных величин x_i .

Определение 1'. Оценка $\hat{\theta}$ называется несмешенной, если для любого возможного значения параметра $\theta \in \Theta$ справедливо равенство

$$\mathbb{E}_\theta \hat{\theta} = \theta, \theta \in \Theta \quad (*).$$

Давайте расшифруем равенство (*):

1) для непрерывной модели оно означает

$$\int \cdots \int \hat{\theta}(x_1, x_2, \dots, x_n) dP_\theta^{(n)}(x_1, x_2, \dots, x_n) = \theta \text{ для всех } \theta,$$

2) для дискретной модели оно означает

$$\sum_i \theta(x_1^i, x_2^i, \dots, x_n^i) dP_\theta^{(n)}(x_1^i, x_2^i, \dots, x_n^i) = \theta \text{ для всех } \theta,$$

Оба определения по-существу эквивалентны. Действительно, как не зная $\theta_{\text{ист}}$ — истинного значения параметра, мы можем проверить, что $\mathbb{E}\hat{\theta} = \theta_{\text{ист}}$. Только проверив справедливость равенства (*) для всех θ , мы будем знать, что (*) будет справедливо и для $\theta_{\text{ист}}$.

Замечание 2. Утверждение о несмещенностии оценки интересно лишь при условии адекватности принятой нами вероятностной модели. Вполне возможна такая ситуация, что справедлива не модель $\{P_\theta\}_{\theta \in \Theta}$, а какая то другая модель $\{\tilde{P}_\theta\}_{\theta \in \Theta}$ с тем же множеством параметров), и данная статистика $\hat{\theta}$ является несмешенной оценкой для модели $\{P_\theta\}$, но не является несмешенной оценкой для модели $\{\tilde{P}_\theta\}_{\theta}$, т.е.

$$\int \cdots \int \hat{\theta}(x_1, x_2, \dots, x_n) d\tilde{P}_\theta^{(n)}(x_1, x_2, \dots, x_n) \neq \theta \text{ для некоторого } \theta.$$

Впрочем, пока не объяснено, как вообще могут существовать такие функции $\hat{\theta}$ (не зависящие от θ !!), что равенство (*) справедливо для всех θ .

Итак, мы рассмотрим примеры несмешенных оценок.

3. Несмешенная оценка среднего.

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

Действительно,

$$\mathbb{E}\bar{x} = \frac{\mathbb{E}x_1 + \mathbb{E}x_2 + \dots + \mathbb{E}x_n}{n} = \frac{n\mathbb{E}x_1}{n} = \mathbb{E}x_1.$$

Итак, — несмешенная оценка среднего случайных величин x_i . Поэтому \bar{x} может служить несмешенной оценкой параметра, если в данной вероятностной модели параметр совпадает со средним распределения. Таким образом, функция является несмешенной оценкой для многих моделей. В частности, в нормальной модели

$$m = \int x dP_{m,\sigma}(x),$$

в модели Бернулли

$$p = \int x dP_p(x),$$

в пуассоновской модели

$$\lambda = \int x dP_\lambda(x)$$

(проверьте это непосредственным вычислением n -кратных интегралов!), таким образом, \bar{x} — несмешенная оценка параметра m в нормальной модели, параметра p в модели Бернулли, параметра λ в модели Пуассона.

4. Несмешенная оценка дисперсии. Сначала вычислим среднее эмпирической дисперсии (не будет ли эмпирическая дисперсия несмешенной оценкой дисперсии так же как эмпирическое среднее \bar{x} — несмешенная оценка среднего). Здесь как и в дальнейшем будет применяться тождество (σ):

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2. \quad (1)$$

Итак, ($m = \mathbb{E}x_i$)

$$\begin{aligned}\mathbb{E}s^2 &= \mathbb{E}\frac{1}{n}\sum_{i=1}^n(x_i - \bar{x})^2 = \mathbb{E}\frac{1}{n}\sum_{i=1}^n[(x_i - m) - (\bar{x} - m)]^2 = \\ &\quad \frac{1}{n}\mathbb{E}\sum_{i=1}^n(x_i - \bar{x})^2 - \mathbb{E}(\bar{x} - m)^2 = \frac{1}{n}n\mathbb{D}x_i - \mathbb{D}\bar{x} = \\ &= \mathbb{D}x_i - \frac{\mathbb{D}x_1 + \mathbb{D}x_2 + \dots + \mathbb{D}x_n}{n^2} = \mathbb{D}x_i - \frac{\mathbb{D}x_i}{n} = \frac{n-1}{n}\mathbb{D}x_i.\end{aligned}$$

Мы видим, что s^2 не является несмешенной оценкой дисперсии, однако несмешенной оценкой дисперсии является величина

$$S^2 = \frac{n}{n-1}s^2. \quad (2)$$

Так как σ^2 — дисперсия нормального распределения, то S^2 — несмешенная оценка квадрата параметра σ в нормальной модели (Отсюда не следует, что $\sqrt{S^2}$ — несмешенная оценка параметра $\sigma!$).

Замечание. Итак, \bar{x} — несмешенная оценка среднего, в любой модели, в которой среднее является параметром, \bar{x} является несмешенной оценкой этого параметра (например, параметра m в нормальной модели, параметра $lambda$ в модели Пуассона и т. д.). Точно так же, $S^2 = \frac{n}{n-1}s^2$ является несмешенной оценкой параметра, если этот параметр является дисперсией в соответствующей модели, например, параметра σ^2 в нормальной модели. Вспомним, что в модели Пуассона параметр λ является не только средним, но и дисперсией. Давайте проверим еще раз это утверждение, но используя характеристическую функцию распределения Пуассона. Напомню сначала, что $\varphi'_\xi(0) = i\mathbb{E}\xi$, $\varphi''_\xi(0) = -\mathbb{E}\xi^2$. Имеем для распределения Пуассона:

$$\varphi_\xi(t) = e^{\lambda(e^{it}-1)}, \varphi'_\xi(t) = i\lambda(e^{it}-1)e^{\lambda(e^{it}-1)}.$$

Поэтому $\varphi'_\xi(0) = i\lambda$, $\varphi''_\xi(0) = -\lambda - \lambda^2$. Отсюда немедленно следует наше утверждение. Итак, какая оценка, \bar{x} или S^2 , лучше для оценки параметра λ .

5. Эффективная оценка.

Определение. Несмешенная оценка $\hat{\theta}$ параметра θ называется *эффективной*, если для любого θ дисперсия $\hat{\theta}$ минимальна в классе всех несмешенных оценок (другими словами, для любой другой несмешенной оценки $\hat{\theta}_1$ параметра θ и любого $\theta \in \Theta$ справедливо неравенство

$$\mathbb{D}_\theta \hat{\theta} \leq \mathbb{D}_\theta \hat{\theta}_1 (\forall \theta \in \Theta). \quad (**)$$

Обсудим это определение. Дисперсия — это 'мера отклонения случайной величины от своего среднего' (см. Курс теории вероятностей), а так как среднее несмешенной оценки совпадает (по ее определению) с истинным значением параметра, то эффективная оценка — это оценка, наименее отклоняющаяся от истинного значения параметра, и это должно быть верно при любом значении параметра.

Замечание. Уменьшение дисперсии оценки без требования несмешенности бессмысленно. Возьмем, например, постоянную оценку $\hat{\theta} \equiv \theta_0$, ее дисперсия минимальна, равна нулю, но что это дает.

Существование эффективных оценок — факт еще более неожиданный, чем существование несмешенных оценок. На первый взгляд, даже построив эффективную оценку, мы не сможем доказать ее эффективность, сравнить ее со всеми несмешенными оценками (как их все рассмотреть). Тем не менее, в некоторых случаях это удается осуществить благодаря неравенству Рао — Крамера.

Теорема. (Неравенство Рао — Крамера). Пусть $\hat{\theta}$ — несмешенная оценка параметра θ . Тогда при выполнении ряда дополнительных условий (которые накладываются по ходу доказательства и читатель может восстановить самостоятельно) справедливо следующее неравенство:

$$\mathbb{D}_{\theta} \hat{\theta} \geq \frac{1}{\mathbb{E}_{\theta} \left[\frac{\partial \ln p_{\theta}^{(n)}}{\partial \theta} \right]^2}. \quad (***)$$

Условия, которые нужны для формулировки (***) очевидны: существование плотности p_{θ} для всех θ , более того, непрерывность и дифференцируемость функции плотности как функции θ при любом фиксированном x , существование дисперсии оценки $\hat{\theta}$ для всех $\theta \in \Theta$. Неравенство Рао — Крамера имеет место и для дискретной модели, когда распределение P_{θ} дискретно для всех θ и сосредоточено на одном и том же для всех θ счетном (или конечном) подмножестве \mathbf{R} . Как и в случае непрерывного типа распределения, вероятность должна дифференцироваться по параметру и должна иметь смысл правая часть в неравенстве Рао — Крамера.

Доказательство теоремы. Из определения несмешенности следует

$$\int \dots \int \hat{\theta}(x_1, x_2, \dots, x_n) p_{\theta}^{(n)}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = \theta, \forall \theta \in \Theta \quad (1)$$

По определению плотности

$$\int \dots \int p_{\theta}^{(n)}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = 1, \forall \theta \in \Theta. \quad (2)$$

Продифференцируем равенства (1) и (2) по θ . Будем считать, что дифференцирование интегралов по параметру θ можно осуществить под знаком интеграла (это еще одно дополнительное условие). Имеем:

$$\int \dots \int \hat{\theta}(x_1, x_2, \dots, x_n) \frac{\partial p_{\theta}^{(n)}(x_1, x_2, \dots, x_n)}{\partial \theta} dx_1 dx_2 \dots dx_n = 1, \forall \theta \in \Theta. \quad (1')$$

$$\int \dots \int \frac{\partial p_{\theta}^{(n)}(x_1, x_2, \dots, x_n)}{\partial \theta} dx_1 dx_2 \dots dx_n = 0, \forall \theta \in \Theta. \quad (2')$$

Вычтем из (1') равенство (2'), умноженное на θ . Мы получили

$$\int \dots \int (\hat{\theta} - \theta) \frac{\partial p_{\theta}^{(n)}(x_1, x_2, \dots, x_n)}{\partial \theta} dx_1 dx_2 \dots dx_n = 1, \forall \theta \in \Theta. \quad (3)$$

Производная плотности не является плотностью вероятности. А в неравенстве Рао — Крамера мы имеем характеристики случайных величин, то есть интегралы по вероятностям. Чтобы привести (3) к такому виду, перепишем равенство (3) следующим образом:

$$\int \dots \int (\hat{\theta} - \theta) \frac{\partial p_{\theta}^{(n)}(x_1, \dots, x_n)}{\partial \theta} \frac{1}{p_{\theta}^{(n)}} p_{\theta}^{(n)} dx_1 \dots dx_n = 1, \forall \theta \in \Theta. \quad (4)$$

или

$$\int \dots \int (\hat{\theta} - \theta) \frac{\partial \ln p_{\theta}^{(n)}}{\partial \theta} p_{\theta}^{(n)} dx_1 \dots dx_n = 1, \forall \theta \in \Theta. \quad (5)$$

Равенство (5) мы интерпретируем как интегрирование по мере $p_{\theta}^{(n)}$ произведения двух функций $(\hat{\theta} - \theta)$ и $\partial \ln(p_{\theta}^{(n)})/\partial \theta$, согласно интегральному варианту неравенства Коши — Буняковского

мы имеем:

$$1 \leq \int \dots \int (\hat{\theta} - \theta)^2 p_{\theta}^{(n)} dx_1 \dots dx_n \int \dots \int \left(\frac{\partial \ln p_{\theta}^{(n)}}{\partial \theta} \right)^2 p_{\theta}^{(n)} dx_1 \dots dx_n = 1,$$

$\forall \theta \in \Theta$, из чего немедленно следует неравенство Рао — Крамера, так как

$$\begin{aligned} \int \dots \int (\hat{\theta} - \theta)^2 p_{\theta}^{(n)} dx_1 \dots dx_n &= \mathbb{D}_{\theta} \hat{\theta}, \\ \int \dots \int \left(\frac{\partial \ln p_{\theta}^{(n)}}{\partial \theta} \right)^2 p_{\theta}^{(n)} dx_1 \dots dx_n &= \mathbb{E}_{\theta} \left[\frac{\partial \ln p_{\theta}^{(n)}}{\partial \theta} \right]^2. \end{aligned}$$

Замечание. Итак, чтобы проверить эффективность некоторой оценки, достаточно проверить для всех θ равенство в неравенстве Рао — Крамера. Сделаем это для оценки \bar{x} параметра m в нормальной модели. Вычислим отдельно левую и правую часть в (***)¹. (По прежнему результаты наблюдений x_i предполагаются независимыми, и это мы сейчас используем для упрощения выкладок в данной задаче (прием бывает полезен и для других моделей и других параметров). А именно, для независимых случайных величин дисперсия суммы равна сумме дисперсий. Поэтому удобно представить среднее квадрата суммы как дисперсию суммы, для чего достаточно проверить равенство нулю среднего суммы.) Имеем

$$\mathbb{D}_{m,\sigma} \bar{x} = \left(\frac{1}{n} \right)^2 (\mathbb{D}x_1 + \mathbb{D}x_2 + \dots + \mathbb{D}x_n) = \frac{\sigma^2 n}{n^2}. \quad (7)$$

$$\ln p_{m,\sigma}^{(n)}(x_1, x_2, \dots, x_n) = \ln \left(\frac{1}{\sigma(2\pi)^{1/2}} \right)^n - \frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - m}{\sigma} \right)^2. \quad (8)$$

Дифференцируя $\ln p_{m,\sigma}^{(n)}$ по m , имеем:

$$\frac{\partial \ln p_{m,\sigma}^{(n)}}{\partial m} = \frac{1}{\sigma^2} \left(\sum_{i=1}^n x_i - nm \right).$$

Очевидно, что

$$\mathbb{E}_{m,\sigma} \frac{1}{\sigma^2} \left(\sum_{i=1}^n x_i - nm \right) = 0.$$

Поэтому

$$\mathbb{E}_{m,\sigma} \left[\frac{\partial \ln p_{m,\sigma}^{(n)}}{\partial m} \right]^2 = \mathbb{D}_{m,\sigma} \left(\frac{1}{\sigma^2} \sum_{i=1}^n x_i \right) = \frac{n\sigma^2}{\sigma^4} = \frac{n}{\sigma^2},$$

что и требовалось.

Упражнение. Докажите эффективность оценки \bar{x} для параметра p в модели Бернулли и для параметра λ в модели Пуассона. Для этого докажите аналог неравенства Рао — Крамера для случая, когда распределение P_{θ} дискретно для всех θ и сосредоточено на одном и том же для всех θ счетном (или конечном) подмножестве \mathbf{R} . Таким, в частности, является случай модели Бернулли и модели Пуассона.

Пример. Далее мы покажем, что оценка \bar{x} не для любой модели является эффективной оценкой среднего. Рассмотрим равномерную модель, в ней число m — середина интервала, на котором сосредоточено равномерное распределение, является средним значением. Действительно,

$$\int_{-\infty}^{\infty} x p_{m,\Delta}(x) dx = \int_{m-\Delta}^{m+\Delta} x \frac{1}{2\Delta} dx = \frac{(m+\Delta)^2 - (m-\Delta)^2}{4\Delta} = m.$$

Тем не менее, \bar{x} не является эффективной оценкой m , значительно меньше дисперсия следующей оценки:

$$\hat{m}(x_1, x_2, \dots, x_n) = \frac{\max x_i + \min x_i}{2}.$$

(Читателю рекомендуется самостоятельно вычислить среднее максимума и минимума и проверить, что оценка \hat{m} — несмещенная оценка параметра m . Впрочем, в этом можно убедиться и используя соображение симметрии. Действительно, преобразование $(x_i) \rightarrow (2m - x_i)$, полученное покоординатным применением преобразования $x \rightarrow 2m - x$, не меняет распределение в пространстве выборок, и не меняет \hat{m} . Следовательно, при этом преобразовании не должно меняться и среднее оценки \hat{m} . Но единственное число, остающееся на месте при преобразовании $x \rightarrow 2m - x$ — это m .)

$$\mathbb{D}\bar{x} = \frac{\sigma^2}{n} = \frac{1}{n} \left(\int_{m-\Delta}^{m+\Delta} x^2 \frac{1}{2\Delta} dx - m^2 \right) = \frac{\Delta^2}{3n}.$$

Далее мы воспользуемся следующим очевидным неравенством:

$$\mathbb{D}(\xi + \eta) \leq 2\mathbb{D}\xi + 2\mathbb{D}\eta. \quad (11)$$

Вычислим дисперсию случайной величины $\max x_i$. Удобнее вместо случайных величин x_i рассмотреть случайные величины $y_i = (x_i - m + \Delta)/2\Delta$, которые имеют равномерное распределение на отрезке $[0, 1]$. Имеем: $F_{\max y_i}(x) = \mathbb{P}\{\max y_i < x\} = \mathbb{P}\{y_1 < x, y_2 < x, \dots, y_n < x\} = x^n$ для $x \in [0, 1]$. Поэтому

$$\begin{aligned} \mathbb{E}[\max y_i] &= \int_0^1 x dx^n = \frac{n}{n+1}, \\ \mathbb{E}[\max y_i]^2 &= \int_0^1 x^2 dx^n = \frac{n}{n+2}, \\ \mathbb{D}[\max y_i] &= \frac{n}{n+2} - \left(\frac{n}{n+1} \right)^2 = \frac{n}{(n+2)(n+1)^2}. \end{aligned}$$

$\mathbb{D}[\max x_i] = O\left(\frac{1}{n^2}\right)$. Ввиду симметрии равномерного распределения $\mathbb{D}[\min x_i] = O\left(\frac{1}{n^2}\right)$,

$$\mathbb{D}[\max x_i + \min x_i]/2 = O\left(\frac{1}{n^2}\right).$$

Итак, для равномерного распределения оценка \hat{m} значительно лучше чем \bar{x} . Отметим, что доказательство неравенства Рао — Крамера в случае равномерной модели не проходит, функция плотности принимает лишь два значения и меняется скачком, т.е. недифференцируема.

6. Оценка максимального правдоподобия. Оценка максимального правдоподобия определяется в случае, когда P_θ имеет плотность распределения (а также в дискретном случае). Эта оценка задается равенством

$$\theta^*(x_1, x_2, \dots, x_n) = \{\theta : p_\theta^{(n)}(x_1, x_2, \dots, x_n) = \max_\theta p_\theta^{(n)}(x_1, x_2, \dots, x_n)\},$$

т.е. в качестве оценки параметра θ берется тот элемент $\theta^*(x_1, \dots, x_n)$ из множества параметров Θ , что для $\theta = \theta^*$ значение плотности

$$p_\theta^{(n)}(x_1, x_2, \dots, x_n) = p_\theta(x_1)p_\theta(x_2) \dots p_\theta(x_n) \quad (***)$$

максимально. Аналогично (вместо плотности берется вероятность точки) определяется оценка максимального правдоподобия в случае, когда используется модель с дискретными распределениями $P_\theta\{x_i\}$, тогда ищется θ , для которого достигается максимум вероятности (а не плотности вероятности).

Замечание. При больших n вероятность, входящая в (***)^{*}, может быть очень мала (в случае модели с плотностью вероятности вероятность точки просто равна нулю), тем не менее, и это удивительно, выбор максимума из очень маленьких вероятностей при разных θ приводит к хорошей оценке:

Теорема. *Если существует эффективная оценка $\hat{\theta}$, для которой достигается равенство в (***) (в условиях неравенства Рао — Крамера), то эта оценка является оценкой максимального правдоподобия.*

Доказательство. В доказательстве неравенства Рао — Крамера используется неравенство Коши — Буняковского, вспомним, что в этом неравенстве равенство достигается лишь в случае пропорциональности двух сомножителей под знаком среднего, итак,

$$\hat{\theta}(x_1, x_2, \dots, x_n) - \theta = \text{Const}(\theta) \frac{\partial \ln p_\theta^{(n)}(x_1, \dots, x_n)}{\partial \theta}$$

для всех $\theta \in \Theta$ и всех x_1, x_2, \dots, x_n .

Подставим в (10) вместо θ оценку максимального правдоподобия $\theta^*(x_1, x_2, \dots, x_n)$ для данных x_1, x_2, \dots, x_n . Так как \ln — монотонная функция, то $\ln p_\theta^{(n)}$ также достигает максимума при $\theta = \theta^*$, но тогда производная

$$\frac{\partial \ln p_\theta^{(n)}(x_1, \dots, x_n)}{\partial \theta} = 0 \text{ при } \theta = \theta^*(x_1, x_2, \dots, x_n),$$

поэтому

$$\hat{\theta}(x_1, x_2, \dots, x_n) = \theta^*(x_1, x_2, \dots, x_n).$$

Алгоритм нахождения эффективной оценки. 1) Найти оценку максимального правдоподобия, приравняв производную по параметру от логарифма функции правдоподобия нулю. 2) Проверить, что эта оценка является непсмешенной. 3) Проверить, что для дисперсии этой оценки и для среднего квадрата производной от логарифма функции правдоподобия выполняется равенство в неравенстве Рао — Крамера.

Пример для нормальной модели.

1) *Оценка параметра m .* Дифференцируем вычисленную выше производную логарифма функции правдоподобия и приравниваем нулю.

$$\frac{1}{\sigma^2} \left(\sum_{i=1}^n x_i - nm \right) = 0.$$

Получаем $m = \bar{x}$. Операции 2) и 3) мы проделали раньше. Заметим, что наш результат не зависит от неизвестного нам параметра σ .

2) *Оценка параметра σ^2 .*

$$\begin{aligned} \frac{\partial \ln p_{m,\sigma}^{(n)}}{\partial \sigma^2} &= -\frac{\partial(n/2) \ln \sigma^2}{\partial \sigma^2} - 1/2 \frac{\partial \sum_{i=1}^n \left(\frac{x_i - m}{\sigma} \right)^2}{\partial \sigma^2} = 0. \\ &-(n/2\sigma^2) + 1/2 \sum_{i=1}^n \frac{(x_i - m)^2}{\sigma^4} = 0. \\ \sigma^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - m)^2. \end{aligned} \tag{*}$$

В этой задаче результат, к сожалению, зависит от неизвестного нам параметра m . Вывод, эффективная оценка может существовать лишь при известном $m = m_0$ и нужно проверить оценку

$$\frac{1}{n} \sum_{i=1}^n (x_i - m_0)^2.$$

Легко проверяется, что среднее этой оценки равно σ^2 (в модели $N(m_0, \sigma)$). Теперь вычислим дисперсию этой оценки.

$$\begin{aligned}\mathbb{D} \left(\frac{1}{n} \sum_{i=1}^n (x_i - m_0)^2 \right) &= \frac{1}{n^2} \sum_{i=1}^n \mathbb{D} ((x_i - m_0)^2) = \\ \frac{n}{n^2} \mathbb{D} ((x_1 - m_0)^2) &= \frac{1}{n} \{ \mathbb{E}(x_1 - m_0)^4 - (\mathbb{E}(x_1 - m_0)^2) \} = \\ \frac{1}{n} \{ 3\sigma^4 - \sigma^4 \} &= \frac{2\sigma^4}{n}.\end{aligned}$$

Теперь считаем знаменатель правой части в неравенстве Рао — Крамера. Производную уже сосчитали. Надо вычислить

$$\mathbb{E} \left(-(n/2\sigma^2) + 1/2 \sum_{i=1}^n \frac{(x_i - m)^2}{\sigma^4} \right)^2.$$

Опять воспользуемся идеей сведения среднего квадрата к дисперсии. Легко проверяется, что

$$\mathbb{E} \left(-(n/2\sigma^2) + 1/2 \sum_{i=1}^n \frac{(x_i - m)^2}{\sigma^4} \right) = 0.$$

Поэтому среднее квадрата равно

$$\mathbb{D} \left(1/2 \sum_{i=1}^n \frac{(x_i - m)^2}{\sigma^4} \right) = n/4 \mathbb{D} \left(\frac{(x_1 - m)^2}{\sigma^4} \right) = \frac{n2\sigma^4}{4\sigma^8}.$$

Еще один пример оценки максимального правдоподобия. Рассматривается равномерная модель $p_\Delta(x) = \frac{1}{\Delta} I_{[0,\Delta]}(x)$. В этой модели равномерное распределение задано на отрезке $[0, \Delta]$, нужно оценить длину отрезка Δ . Ясно, что если результаты наблюдения — числа x_1, x_2, \dots, x_n , то $\Delta \geq \max\{x_i\}$. Оценка максимального правдоподобия равна $\max\{x_i\}$, так как при $\Delta = \max\{x_i\}$ значение функции

$$p_\Delta^{(n)}(x_1, \dots, x_n) = \left(\frac{1}{\Delta} \right)^n \prod_i I_{[0,\Delta]}(x_i)$$

максимально и равно $\left(\frac{1}{\max\{x_i\}} \right)^n$.

7. Состоятельная оценка.

Во многих задачах оценка как функция n переменных задается формулой, имеющей смысл для любого n . Более того, в сколь-нибудь общей теории оценок бессмысленно строить ее лишь для некоторого фиксированного числа наблюдений. Итак, мы имеем полное право трактовать оценку не как фиксированную функцию на \mathbf{R}^n , а как набор функций $\hat{\theta}^{(n)}$, каждая из которых задана на своем пространстве выборок \mathbf{R}^n , при этой трактовке каждому n соответствует некоторое распределение $P_{\theta, \hat{\theta}^{(n)}}$ (заданное на множестве Θ). Вот почему имеет смысл следующее определение:

Определение. Оценка $\hat{\theta}^{(n)}(n \rightarrow \infty)$ называется *состоятельной*, если $\hat{\theta}^{(n)} \xrightarrow{\mathbb{P}} \theta$ (для любого θ) или, что то же (см. теорему о совпадении слабой сходимости и сходимости по вероятности в случае постоянства предельной случайной величины), если

$$P_{\theta, \hat{\theta}^{(n)}} \xrightarrow{\text{СЛ}} \delta_\theta \forall \theta \in \Theta,$$

где δ_θ — вероятность, сосредоточенная в точке θ , $\delta_\theta\{\theta\} = 1$. Заметим, что в терминах статистической модели мы должны предположить стремление размерности пространства выборок

(например, \mathbf{R}^n) к ∞ , тогда говорить о сходимости случайных величин $\hat{\theta}^{(n)}$ некорректно, так как эти случайные величины заданы на разных вероятностных пространствах. Но говорить о слабой сходимости распределений этих оценок мы имеем право, так как эти распределения заданы на одном пространстве параметров.

Теорема. *Оценка \bar{x} является состоятельной оценкой параметра θ нормального распределения.*

Доказательство. По закону больших чисел

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}x_i = m.$$

Замечание. И в других моделях, где параметр θ является средним распределения с параметром θ , \bar{x} является состоятельной оценкой параметра θ (в частности, \bar{x} — состоятельная оценка параметра p в модели Бернулли и параметра λ в модели Пуассона).

Упражнение 1 (обязательное). Разобраться в примечании. Из какой теоремы это следует?

Упражнение 2 (обязательное). Доказать, что S^2 — состоятельная оценка дисперсии σ^2 в нормальной модели. [Указание. Если ξ имеет нормальное распределение с параметрами 0 и σ , то $\mathbb{E}\xi^4 = 3!\sigma^4 = 3\sigma^4$.] Будет ли состоятельной оценкой дисперсии в нормальной модели оценка s^2 ? Можно ли утверждать, что S^2 является состоятельной оценкой параметра θ , если этот параметр совпадает с дисперсией в нашей модели?

Упражнение 3. Рассмотрим модель Коши, заданную набором распределений с плотностями $p_\mu(x) = \frac{a}{\pi[a^2(x - \mu)^2 + 1]}$, $a > 0$. Доказать, что \bar{x} не является состоятельной оценкой параметра μ .

[Указание. Используя характеристическую функцию распределения Коши, вычислить характеристическую функцию распределения случайной величины \bar{x} , доказать, что это распределение не зависит от n .] [Решение. Искомая характеристическая имеет вид $\varphi_{x_k}(t) = e^{ia\mu t - |at|}$ для всех k . Поэтому

$$\varphi_{\bar{x}}(t) = (\varphi_{x_1}(t/n))^n = \varphi_{x_1}(t),$$

то есть никакой сходимости к константе мы не имеем.] Разумеется, в этой модели параметр μ не является средним, но находится посередине распределения, является медианой. Покажите, что медиана эмпирического распределения сходится к μ .

Следующая теорема указывает общий путь получения состоятельных оценок в нетривиальных ситуациях.

8. Теорема. *Оценка максимального правдоподобия является состоятельной оценкой. Мы будем считать, что распределения P_θ нашей модели имеют непрерывный тип.*

При этом на модель $\{P_\theta\}_{\theta \in \Theta}$ накладываются следующие условия (может их не помнить, они сами возникнут):

1) *плотности p_θ (задаваемые с точностью до множества лебеговой меры нуль) могут быть заданы так, что множества $\{p_\theta \neq 0\}$ не зависят от θ ;*

2) *при всех возможных выборках $\{x_1, x_2, \dots, x_n\}$ функция $\theta \rightarrow p_\theta^{(n)}(x_1, \dots, x_n)$ имеет единственный локальный максимум.*

Этот максимум, который мы обозначим $\hat{\theta}$, и называется (напомним это) оценкой максимального правдоподобия.

Заметим, что нарушение условия 1) может даже облегчить ситуацию — если результат наблюдения x_i таков, что $x_i \in (a, b)$ и $\int_a^b p_\theta(x)dx = 0$, то мы вправе исключить данное θ из претендентов на нашу оценку.

Итак, нам нужно доказать, что

$$\mathbb{P}_\theta^{(n)}\{(x_1, \dots, x_n) : \hat{\theta} \in (\theta - \varepsilon, \theta + \varepsilon)\} \rightarrow 1 \quad \forall \varepsilon > 0 \forall \theta.$$

Введенное выше условие позволяет заменить событие в формуле на меньшее:

$$\{\hat{\theta} \in (\theta - \varepsilon, \theta + \varepsilon)\} \supset \{(x_1, \dots, x_n) : p_\theta^{(n)}(x_1, \dots, x_n) > p_{\theta-\varepsilon}^{(n)}(x_1, \dots, x_n),$$

$$p_\theta^{(n)}(x_1, \dots, x_n) > p_{\theta+\varepsilon}^{(n)}(x_1, \dots, x_n)\}. \quad (*)$$

Действительно, если значения на границе интервала строго меньше значений на середине, то в данном интервале имеется локальный максимум, который, согласно допущению, является оценкой максимального правдоподобия. Обратная импликация, вообще говоря, неверна.

Итак, достаточно доказать две сходимости:

$$\mathbb{P}_\theta^{(n)}\{(x_1, \dots, x_n) : p_\theta^{(n)}(x_1, \dots, x_n) > p_{\theta-\varepsilon}^{(n)}(x_1, \dots, x_n)\} \rightarrow 1,$$

$$\mathbb{P}_\theta^{(n)}\{(x_1, \dots, x_n) : p_\theta^{(n)}(x_1, \dots, x_n) > p_{\theta+\varepsilon}^{(n)}(x_1, \dots, x_n)\} \rightarrow 1.$$

Оба соотношения доказываются одинаково, мы докажем только второе. Запишем событие другим способом:

$$\begin{aligned} \{p_\theta^{(n)} > p_{\theta+\varepsilon}^{(n)}\} &= \left\{ \ln \left(\frac{p_{\theta+\varepsilon}^{(n)}}{p_\theta^{(n)}} \right) < 0 \right\} = \\ &= \left\{ (x_1, \dots, x_n) : \frac{1}{n} \sum_{i=1}^n \ln \left(\frac{p_{\theta+\varepsilon}(x_i)}{p_\theta(x_i)} \right) < 0 \right\}. \end{aligned}$$

Это напоминает закон больших чисел. Действительно, случайные величины $\ln \left(\frac{p_{\theta+\varepsilon}(x_i)}{p_\theta(x_i)} \right)$ (теперь x_i — это случайные величины, а не результаты наблюдений) независимы как функции независимых случайных величин x_i . Поэтому мы левую часть определяющего событие соотношения заранее разделили на n . Предположим дополнительно, что $\mathbb{E}_\theta \ln \frac{p_{\theta+\varepsilon}(x_i)}{p_\theta(x_i)}$ существует. Если мы докажем, что это среднее строго отрицательно, то из слабого закона больших чисел (но в условиях Колмогорова, в предположении существования лишь среднего, см. раздел 'Характеристические функции') будет следовать, что $\frac{1}{n} \sum_{i=1}^n \ln \frac{p_{\theta+\varepsilon}(x_i)}{p_\theta(x_i)}$ сходится по вероятности к строго отрицательному числу. Но тогда при $n \rightarrow \infty$ вероятность интересующегося нас события будет сходиться к 1. Переходим к доказательству нужного неравенства

$$\mathbb{E}_\theta \ln \left(\frac{p_{\theta+\varepsilon}(x_i)}{p_\theta(x_i)} \right) < 0.$$

Мы воспользуемся нужным нам и в дальнейшем неравенством выпуклости для среднего.

Лемма. (Неравенство для выпуклых функций.) *Если g строго выпуклая функция, то для любой случайной величины ξ , для которой правая и левая части в (1) имеют смысл, имеет место неравенство*

$$\mathbb{E}g(\xi) \geq g(\mathbb{E}\xi), \quad (1)$$

причем равенство достигается лишь в случае константы ξ .

Мы не будем давать здесь подробного доказательства, но постараемся объяснить, почему неравенство (1) верно.

Оно верно в случае двузначной случайной величины ξ : $\mathbb{P}_\xi\{x_1\} = p$, $\mathbb{P}_\xi\{x_2\} = 1 - p$:

$$\mathbb{E}g(\xi) = pg(x_1) + (1-p)g(x_2) \geq g(px_1 + (1-p)x_2) = g(\mathbb{E}\xi),$$

где неравенство является определением выпуклости. Методом математической индукции легко доказывается неравенство для ξ со значениями в любом конечном множестве $\{x_1, \dots, x_n\}$:

$$p_1g(x_1) + \dots + p_ng(x_n) \geq g(p_1x_1 + \dots + p_nx_n), p_i \geq 0, \sum_i p_i = 1.$$

Переходом к пределу и использованием непрерывности получается неравенство для бесконечных сумм

$$\sum_{i=1}^{\infty} p_i g(x_i) \geq g \left(\sum_{i=1}^{\infty} p_i x_i \right), p_i \geq 0, \sum_i p_i = 1. \quad (2)$$

Последнее неравенство означает (1) для дискретных случайных величин, для которых все ряды сходятся. Так как любая случайная величина является равномерным пределом дискретных случайных величин, переходом к пределу в (2) для случайных величин, мы получаем (1) в общем случае.

Заметим, что во всех этих равенствах при строгой выпуклости g , ненулевых значениях p_i и различных x_i и равенства быть не может. Верно и более общее утверждение, что равенство невозможно при строгой выпуклости g и случайной величине ξ , не равной константе.

Можно убедиться в справедливости неравенства выпуклости и из геометрических соображений. Для этого удобно рассмотреть случайный вектор $(\xi, g(\xi))$. Распределение этого вектора лежит на выпуклой кривой $\{(x, y) : y = g(x)\}$, а среднее $(\mathbb{E}\xi, \mathbb{E}g(\xi))$ находится в выпуклой области, лежащей над этой кривой, это очевидно из механических соображений, среднее является центром тяжести. Но тогда $\mathbb{E}g(\xi)$ будет выше на графике чем $g(\mathbb{E}\xi)$.

Замечание. Для строго вогнутой функции \ln из леммы следует справедливость обратного неравенства:

$$\begin{aligned} \mathbb{E}_\theta \ln \left(\frac{p_{\theta+\varepsilon}(x_i)}{p_\theta(x_i)} \right) &< \ln \left(\mathbb{E}_\theta \left[\frac{p_{\theta+\varepsilon}(x_i)}{p_\theta(x_i)} \right] \right) = \\ &= \ln \left(\int_{\mathbb{R}} \frac{p_{\theta+\varepsilon}(x)}{p_\theta(x)} p_\theta(x) dx \right) = \ln \left(\int_{\mathbb{R}} p_{\theta+\varepsilon}(x) dx \right) = \ln 1 = 0. \end{aligned}$$

Это завершает доказательство теоремы в предположении существования $\mathbb{E}_\theta \ln \left(\frac{p_{\theta+\varepsilon}(x_i)}{p_\theta(x_i)} \right)$.

Но что делать, когда $\mathbb{E}_\theta \ln \left(\frac{p_{\theta+\varepsilon}(x_i)}{p_\theta(x_i)} \right)$ не существует. В этом случае из условия 1) следует, что

$$P_\theta \left\{ x : -\infty < \ln \left(\frac{p_{\theta+\varepsilon}(x)}{p_\theta(x)} \right) < +\infty \right\} = 1.$$

Сейчас мы это используем.

Любую случайную величину ξ мы можем представить в виде суммы неположительной и неотрицательной случайных величин:

$$\xi = \xi^+ + \xi^-, \text{ где } \xi^+ = \max\{\xi, 0\}, \xi^- = \min\{\xi, 0\}.$$

Случайная величина ξ интегрируема тогда и только тогда, когда $\mathbb{E}\xi^+ < \infty$, $\mathbb{E}\xi^- > -\infty$. Но возможны другие варианты: $\mathbb{E}\xi^+ < \infty$, $\mathbb{E}\xi^-$ неинтегрируема, тогда мы считаем, что $\mathbb{E}\xi = -\infty$, или $\mathbb{E}\xi^- > -\infty$, $\mathbb{E}\xi^+$ неинтегрируема, тогда мы считаем, что $\mathbb{E}\xi = \infty$. Но $\mathbb{E}\xi$ не имеет никакого смысла, если оба интеграла не существуют. (Заметим, что мы только что перебрали все возможные варианты.) Мы должны доказать, что либо $\mathbb{E}_\theta \left(\ln \left[\frac{p_{\theta+\varepsilon}(x)}{p_\theta(x)} \right] \right)$ существует, либо

$$\mathbb{E}_\theta \left(\ln \left[\frac{p_{\theta+\varepsilon}(x)}{p_\theta(x)} \right] \right)^+ < \infty. \quad (3)$$

Так как наша случайная величина не принимает значений $\pm\infty$, нам достаточно доказать, что для любых $C_n \nearrow \infty$, $D_n \searrow -\infty$,

$$\limsup_{n \rightarrow \infty} \int_{D_n < \ln \left[\frac{p_{\theta+\varepsilon}(x)}{p_\theta(x)} \right] < C_n} \ln \left(\frac{p_{\theta+\varepsilon}(x)}{p_\theta(x)} \right) p_\theta(x) dx < \infty. \quad (4)$$

(Если мы докажем (4), то (3) не может не выполняться. Действительно, если $D < 0$, то число

$$-\int_{D < \ln \left[\frac{p_{\theta+\varepsilon}(x)}{p_\theta(x)} \right] < 0} \ln \left(\frac{p_{\theta+\varepsilon}(x)}{p_\theta(x)} \right) p_\theta(x) dx$$

в принципе может быть сколь угодно большим для больших $|D|$, но это число всегда конечно (интеграл по вероятностной мере по функции, которая не превосходит $|D|$). В то же время, если неверно (3) и интеграл равен ∞ , то

$$\lim_{C \rightarrow \infty} \int_{0 < \ln \left[\frac{p_{\theta+\varepsilon}(x)}{p_\theta(x)} \right] < C} \ln \left(\frac{p_{\theta+\varepsilon}(x)}{p_\theta(x)} \right) p_\theta(x) dx = \infty.$$

Поэтому по любой последовательности $D_n \searrow -\infty$ можно выбрать такую быстро сходящуюся последовательность $C_n \nearrow \infty$, что

$$- \int_{D_n < \ln \left[\frac{p_{\theta+\varepsilon}(x)}{p_\theta(x)} \right] < 0} \ln \left(\frac{p_{\theta+\varepsilon}(x)}{p_\theta(x)} \right) p_\theta(x) dx < \int_{0 < \ln \left[\frac{p_{\theta+\varepsilon}(x)}{p_\theta(x)} \right] < C} \ln \left(\frac{p_{\theta+\varepsilon}(x)}{p_\theta(x)} \right) p_\theta(x) dx + n$$

и тогда (4) будет неверно.) Дальше мы будем считать последовательности $C_n \nearrow \infty$, $D_n \searrow -\infty$ произвольными и для упрощения обозначений мы введем последовательность множеств

$$A_n = \left\{ x : D_n < \ln \left[\frac{p_{\theta+\varepsilon}(x)}{p_\theta(x)} \right] < C_n \right\}.$$

На множестве A_n функции p_θ и $p_{\theta+\varepsilon}$ не обязаны быть плотностями, эту ситуацию нужно исправить. Обозначим

$$\int_{A_n} p_\theta(x) dx = 1 - \lambda_n, \quad \int_{A_n} p_{\theta+\varepsilon}(x) dx = 1 - \mu_n,$$

и пронормируем функции так, что они станут функциями плотности на A_n :

$$\tilde{p}_\theta(x) = \frac{p_\theta(x)}{1 - \lambda_n}, \quad \tilde{p}_{\theta+\varepsilon}(x) = \frac{p_{\theta+\varepsilon}(x)}{1 - \mu_n}.$$

Очевидно, что $\lambda_n \rightarrow 0$, $\mu_n \rightarrow 0$. Как и выше, доказывается, что

$$\int_{A_n} \ln \left(\frac{\tilde{p}_{\theta+\varepsilon}(x)}{\tilde{p}_\theta(x)} \right) \tilde{p}_\theta(x) dx = \tilde{\pi}_n \leq 0.$$

В то же время

$$\begin{aligned} \pi_n &= \int_{A_n} \ln \left(\frac{p_{\theta+\varepsilon}(x)}{p_\theta(x)} \right) p_\theta(x) dx = \\ &= \int_{A_n} \ln \left(\frac{(1 - \mu_n) \tilde{p}_{\theta+\varepsilon}(x)}{(1 - \lambda_n) \tilde{p}_\theta(x)} \right) (1 - \lambda_n) \tilde{p}_\theta(x) dx = \\ &= \int_{A_n} \left\{ [\ln(1 - \mu_n) - \ln(1 - \lambda_n)] + \ln \left(\frac{\tilde{p}_{\theta+\varepsilon}(x)}{\tilde{p}_\theta(x)} \right) \right\} (1 - \lambda_n) \tilde{p}_\theta(x) dx = \\ &= \tilde{\pi}_n (1 - \lambda_n) + [\ln(1 - \mu_n) - \ln(1 - \lambda_n)] \int_{A_n} p_\theta(x) dx. \end{aligned}$$

Итак, π_n никак не может сходиться к $+\infty$, так как $\tilde{\pi}_n$ неположительна, $(1 - \lambda_n)$ неотрицательна, разность логарифмов сходится к 0, а интеграл ограничен 1.

Завершая наши рассуждения, заметим, что закон больших чисел Колмогорова для независимых одинаково распределенных случайных величин ξ_n легко переписывается на случай среднего значения ξ_n , равного $-\infty$ (тогда интеграл от положительной части ξ_n — фиксированное положительное число). Действительно, в этом случае мы представляем $\xi_n = \eta_n + \zeta_n$, где новые случайные величины также одинаково распределены и независимы при разных n , $\zeta_n < 0$, а случайные величины η_n имеют среднее, но это среднее может быть сделано (при соответствующем выборе разложения ξ_n) сколь угодно маленьким (меньше $-N$). Чтобы этого достигнуть, мы можем добавить к положительной части ξ достаточно большую долю отрицательной части (это возможно, так как интеграл от отрицательной части сходится к $-\infty$).

С вероятностью 1

$$\limsup_n \frac{\xi_1 + \dots + \xi_n}{n} \leq \lim_n \frac{\eta_1 + \dots + \eta_n}{n} \leq -N.$$

для всех N . Итак, с вероятностью 1 усредненные суммы для ξ_i сходятся к $-\infty$.

Комментарий. В завершение параграфа я объясню целесообразность введения понятия состоятельной оценки. Рассмотрим, для примера, нормальную модель. \bar{x} — эффективная оценка параметра m . Но в то же время утверждать, что \bar{x} точно равно m , мы можем лишь с нулевой вероятностью. Это следует из непрерывности распределения \bar{x} . Практикам хочется большей определенности, пусть даже за счет отказа от точности нашей оценки. Немного размыв нашу оценку, мы можем заменить неприятное утверждение

$$\mathbb{P}\{m = \bar{x}\} = 0$$

на весьма привлекательное утверждение

$$\mathbb{P}\{m \in (\bar{x} - \varepsilon, \bar{x} + \varepsilon)\} \rightarrow 1.$$

При больших n мы можем доверять утверждению $m \in (\bar{x} - \varepsilon, \bar{x} + \varepsilon)$ с вероятностью, близкой к 1. Это подводит нас к понятию доверительного интервала.

Вопрос. Выполняются ли условия теоремы в модели Коши? Очевидно, что в случае лишь одного наблюдения выполняются. Но что будет в многомерном случае? Рассмотрите случай двух наблюдений.

Замечание. Простой пример, в котором условия теоремы не выполняются, это — равномерная модель с фиксированной длиной интервала и меняющейся серединой интервала m . Ясно, что множества нулевой плотности для разных m разные, поэтому 1) неверно. Впрочем, и 2) также не выполняется.

9. Некоторые универсальные оценки параметра θ

В этом пункте мы рассмотрим некоторые алгоритмы оценивания параметра, применимые сразу для многих или даже для всех моделей.

Оценка максимального правдоподобия.

Метод моментов.

Метод минимума χ^2 .

Группировка данных и поправки Шеппарда.

§19. Доверительные интервалы

В прошлой лекции было введено понятие оценки, рассматривались наилучшие оценки, тем не менее, даже самые хорошие оценки параметра для практиков не всегда приемлемы. Действительно, им недостаточно утверждения ' $\hat{\theta}$ близко к истинному значению параметра θ' ', а хотелось бы утверждение типа ' $\hat{\theta}$ равно истинному значению параметра θ' ' что, как правило, неверно, а в случае модели с непрерывным множеством значений параметра θ , равенство $\hat{\theta} = \theta_{\text{ист}}$ может выполняться только с вероятностью 0 (проверьте это для нормальной модели). Поэтому практики готовы пожертвовать точностью оценки, придать ей более расплывчатый характер, но зато усилить достоверность этой расплывчатой оценки.

1. Определение. Доверительным интервалом для параметра (в модели $\{P_\theta\}$, $\theta \in \Theta$) служит интервал

$$(\hat{\theta}_1(x_1, x_2, \dots, x_n), \hat{\theta}_2(x_1, x_2, \dots, x_n)),$$

задаваемый двумя функциями $\hat{\theta}_1$ и $\hat{\theta}_2$ от результатов наблюдений и такой, что

$$\inf_{\theta \in \Theta} \mathbb{P}_\theta^{(n)} \{(x_i) : (\hat{\theta}_1(x_1, x_2, \dots, x_n), \hat{\theta}_2(x_1, x_2, \dots, x_n))\} \geq \gamma, \quad (*)$$

здесь γ — доверительный уровень — число между нулем и единицей (разумеется, близкое к 1), которое задается заказчиком вашей статистической разработки.

Замечание 1. Как и в определении несмещенной или эффективной оценки, мы добиваемся цели — достоверности оценки для вероятности при неизвестном нам значении параметра $\theta = \theta_{\text{ист}}$ благодаря требованию справедливости этого неравенства для всех возможных значений параметра.

Замечание 2. Для некоторых случаев удается построить такой доверительный интервал $(\hat{\theta}_1, \hat{\theta}_2)$, что тождественно для все $\theta \in \Theta$

$$\mathbb{P}_\theta^{(n)} \{(x_i) : (\hat{\theta}_1(x_1, x_2, \dots, x_n), \hat{\theta}_2(x_1, x_2, \dots, x_n))\} = \gamma, \quad (**)$$

Замечание 3. Обычно в основе определения доверительного интервала служит некоторая статистика — функция T на пространстве выборок \mathbf{R}^n такая, что распределение некоторого простого преобразования T , использующего параметр θ (обычно это сложение, деление, умножение или вычитание) не зависит от неизвестного нам значения параметра θ . Как правило, распределение этой величины $f(T, \theta)$ отражено в математических таблицах. Итак, если распределение $f(T, \theta)$ не зависит от θ и равно \mathbb{Q} , то доверительный интервал строится следующим образом: сначала выбираются такие a_1 и a_2 , что

$$\mathbb{P}_\theta^{(n)} \{(x_i) : a_1 < f(T, \theta) < a_2\} = \mathbb{Q}\{x : a_1 < x < a_2\} = \gamma \forall \theta \in \Theta,$$

далее неравенства $a_1 < f(T, \theta) < a_2$ решаются относительно θ . Можно сказать так, что сначала доверительный интервал выбирается на области значений $f(T, \theta)$, в результате чего достигается равенство а после простыми преобразованиями переносится на область значений параметров. Заметим также, что в рассмотренных нами ниже задачах для нормальной модели $N(m, \sigma)$ распределение $f(T, \theta)$ будет стандартным нормальным, распределением Стьюдента или χ^2 . Сначала мы покажем, что $f(T, \theta)$ — функция независимых стандартных нормальных случайных величин $\frac{x_i - m}{\sigma}$, в потом докажем, что ее распределение является табличным.

2. Доверительные интервалы в нормальной модели.

Напомним, что нормальная вероятностная модель задается набором распределений с функцией плотности

$$p_{m, \sigma}(x) = \frac{1}{\sigma(2\pi)^{1/2}} \exp\left(-\frac{1}{2} \left(\frac{x - m}{\sigma}\right)^2\right), x \in \mathbf{R}, \text{ где } m \in \mathbf{R}, \sigma \geq 0,$$

нормальная статистическая модель задается набором n -мерных распределений с функцией плотности

$$p_{m, \sigma}^{(n)}(x_1, x_2, \dots, x_n) = \left(\frac{1}{\sigma(2\pi)^{1/2}}\right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - m}{\sigma}\right)^2\right), x_i \in \mathbf{R},$$

Здесь m пробегает всю числовую прямую, σ — множество всех положительных чисел.

Пример. Хорошо известно, что результат измерения расстояния имеет нормальное распределение, при этом, если нет систематической ошибки измерения, то m — это искомое расстояние, σ характеризует ошибку измерения. Обычно интересует либо то, либо другое: либо мы

хотим знать расстояние (и для нас несущественна точность прибора, ее мы определять не хотим, хотя она конечно должна повлиять на длину доверительного интервала для m), если нас интересует прибор, то, конечно же, нам не важно расстояние, мы занимаемся только исследованием прибора. Итак, возникают четыре статистические задачи при оценивании параметров нормального распределения: 1) оценка среднего при известной дисперсии, 2) оценка среднего при неизвестной дисперсии, 3) оценка дисперсии при известном среднем, 4) оценка дисперсии при неизвестном среднем.,

Задача 1. Оценка среднего при известной дисперсии σ_0^2 .

Так как x_i — независимые нормальные $N(m, \sigma_0)$ случайные величины, то $x_1 + x_2 + \dots + x_n$ нормальна $N(nm, n^{1/2}\sigma_0)$ (см. устойчивость нормального распределения в разделе характеристическая функция), поэтому \bar{x} имеет параметры $N(m, n^{-1/2}\sigma_0)$, случайная величина $n^{1/2}(\bar{x} - m)/\sigma_0$ имеет распределение стандартное $N(0, 1)$. Пусть теперь нам нужно построить доверительный интервал для m с доверительным уровнем $\gamma = 0.9$. Ищем такое ε_γ , что для случайной величины $\xi = n^{1/2}(\bar{x} - m)/\sigma_0$, имеющей стандартное распределение $N(0, 1)$, справедливо равенство

$$\mathbb{P}\{|\xi| \leq \varepsilon_\gamma\} = 0.9.$$

Так как распределение ξ симметрично, то $\mathbb{P}\{\xi > t\} = \mathbb{P}\{\xi < -t\}$ для всех t , находим по таблице стандартного нормального распределения (функция распределения обычно обозначается через Φ) такое ε_γ , что

$$\Phi(\varepsilon_\gamma) = \int_{-\infty}^{\varepsilon_\gamma} \frac{1}{(2\pi)^{1/2}} \exp(-\frac{1}{2}x^2) dx = 0.95,$$

тогда

$$\mathbb{P}\{|\xi| \leq \varepsilon_\gamma\} = \int_{-\varepsilon_\gamma}^{\varepsilon_\gamma} \frac{1}{(2\pi)^{1/2}} \exp(-\frac{1}{2}x^2) dx = 0.9.$$

Легко видеть, что $\varepsilon_\gamma = 1.65$. Итак,

$$\mathbb{P}\{|n^{1/2}(\bar{x} - m)/\sigma_0| \leq 1.65\} = 0.9.$$

Разрешая это неравенство относительно m , получаем доверительный интервал для m : $m \in (\bar{x} - (\sigma_0/n^{1/2})1.65, \bar{x} + (\sigma_0/n^{1/2})1.65)$.

Замечание 1. Иногда представляет интерес доверительный интервал, у которого один конец равен $+\infty$ или $-\infty$. В данном случае доверительный интервал $(-\infty, \bar{x} + (\sigma_0/n^{1/2})1.65)$ будет давать доверительный уровень не 0.9, а 0.95.

Замечание 2. Иногда заказчик может сам задать длину доверительного интервала и доверительный уровень. Вы тогда, зная σ_0 , можете определить необходимое число наблюдений.

Пример. При определении глубины моря в заданной точке $\sigma_0 = 20$ м, требуемый диаметр доверительного интервала равен 30 м (радиус = 15м), определить количество наблюдений, необходимых для построения доверительного интервала заданной длины и заданным доверительным уровнем $\gamma = 0.9$.

Решение. Решаем неравенство относительно n :

$$20n^{-1/2}1.65 \leq 15, n^{1/2} \geq 2.2, n \geq 4.84.$$

Хотя это вряд ли разумно, можно последнее 5-е наблюдение проводить с вероятностью 0,84 (или не проводить с вероятностью 0,16), используя для принятия решения о проведении наблюдения датчик случайных чисел.

Задача 2. Оценка дисперсии σ^2 при известном среднем m_0 .

В этой задаче рассматривается статистика $S_{m_0}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - m_0)^2$ (это новое обозначение).

Мы так преобразуем случайную величину $S_{m_0}^2$, чтобы получилась новая, распределение которой не зависит от σ и более того, является табличным распределением χ_n^2 . Действительно, мы

покажем, что случайная величина

$$T(x_1, x_2, \dots, x_n) = \frac{nS_{m_0}^2}{\sigma^2}$$

имеет распределение χ_n^2 . Напомним, что распределение χ_n^2 — это распределение случайной величины $\xi_1^2 + \xi_2^2 + \dots + \xi_n^2$, где $\xi_1, \xi_2, \dots, \xi_n$ — независимые одинаково распределенные нормальные $N(0, 1)$ случайные величины. Имеем:

$$\frac{nS_{m_0}^2}{\sigma^2} = \sum_{i=1}^n (x_i - m_0)^2 / \sigma^2$$

представляется в виде суммы χ^2 , где $\xi_i = (x_i - m_0) / \sigma$.

Используя таблицы распределения случайной величины χ_n^2 , строится доверительный интервал σ^2 . Чаще всего доверительный интервал строится в виде (α, β) (если нас интересует наиболее точное задание дисперсии), $(0, \delta)$ (если нас интересует вопрос — может ли дисперсия оказаться больше какого-то числа, быть слишком большой), и (τ, ∞) (если нас интересует — может ли дисперсия оказаться слишком маленькой). Доверительный интервал всех трех видов строится одинаковым способом. Построим, например, доверительный интервал третьего вида: выберем такое $\lambda > 0$, что

$$\mathbb{P} \left\{ \frac{nS_{m_0}^2}{\sigma^2} < \lambda \right\} = \gamma.$$

Мы получаем, что интервал $\left(\frac{nS_{m_0}^2}{\lambda}, \infty \right)$ имеет доверительный уровень

$$\mathbb{P} \left\{ \frac{nS_{m_0}^2}{\lambda} < \sigma^2 \right\} = \gamma.$$

Замечание 3. Как видно из задачи 1, знание дисперсии позволяет оценить точность оценки \bar{x} . Тогда доверительный интервал для дисперсии — это оценка точности точности. Интуитивно ясно, что для такой оценки нужно существенно больше наблюдений чем в задаче 1. Однако для больших n таблиц распределения χ_n^2 не существует. Нужно пользоваться нормальным приближением — случайная величина $\frac{\chi_n^2 - n}{\sqrt{2n}}$ имеет при больших n почти стандартное нормальное распределение. В таких ситуациях нужно (умеючи!) пользоваться таблицами нормального распределения $N(0, 1)$.

Задача 3. Оценка дисперсии σ^2 при неизвестном среднем.

В этом случае мы не можем использовать центрирование случайных величин x_i средним m_0 , поэтому вместо статистики $S_{m_0}^2$ здесь используется несмещенная оценка дисперсии

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Теорема. Случайная величина

$$(n-1) \frac{S^2}{\sigma^2} = \sum_{i=1}^n (x_i - \bar{x})^2 / \sigma^2$$

имеет распределение χ_{n-1}^2 .

Доказательство. Мы используем тождество (1) в §2, а также следующую лемму о вращении нормального случайного вектора:

Лемма 1. Пусть $(\xi_1, \xi_2, \dots, \xi_n)$ — стандартный нормальный случайный вектор (т.е. распределения $\xi_1, \xi_2, \dots, \xi_n$ нормальны $N(0, 1)$, сами эти величины независимы). Пусть также U —

ортогональное вращением, $U = [u_{ij}]$, $i, j \leq n$, — матрица вращения, случайный вектор (η_i) задан равенствами

$$\eta_i = \sum_{i=1}^n u_{ij} \xi_j, i \leq n.$$

Тогда случайный вектор $(\eta_1, \eta_2, \dots, \eta_n)$ также имеет стандартное n -мерное нормальное распределение.

Вернемся к доказательству теоремы. Обозначим $\xi_i = \frac{x_i - m}{\sigma}$, $\bar{\xi} = \frac{\xi_1 + \xi_2 + \dots + \xi_n}{n}$. Тогда

$$\begin{aligned} T(x_1, x_2, \dots, x_n) &= \sum_{i=1}^n (x_i - \bar{x})^2 / \sigma^2 = \sum_{i=1}^n [(x_i - m) - (\bar{x} - m)]^2 / \sigma^2 = \\ &= \sum_{i=1}^n (\xi_i - \bar{\xi})^2 = \sum_{i=1}^n \xi_i^2 - n\bar{\xi}^2. \end{aligned}$$

Далее мы построим такое вращение U , матрица \mathbf{U} которого имеет вид

$$\mathbf{U} = \begin{bmatrix} 1/\sqrt{n} & \dots & 1/\sqrt{n} \\ \dots & \dots & \dots \\ \dots & \dots & \dots \end{bmatrix}.$$

Задана 1-я строка, в последующих строках стоят какие-то числа достаточно произвольно, но так, чтобы матрица \mathbf{U} была ортогональной (Напомним, строки и столбцы ортогональной матрицы составляют ортонормированный базис. Любую ортонормированную систему, в частности, даже один нормированный вектор, можно дополнить до ортонормированного базиса).

В этом случае $\eta_1 = n^{-1/2}(\xi_1 + \dots + \xi_n)$. Так как при ортогональном вращении сумма квадратов переходит в сумму квадратов, то $\sum_{i=1}^n \xi_i^2 - n\bar{\xi}^2 = \sum_{i=1}^n \eta_i^2 - \eta_1^2 = \sum_{i=2}^n \eta_i^2$, т.е. имеет распределение χ_{n-1}^2 .

Задача 4. Оценка среднего при неизвестной дисперсии.

При решении задачи 1 мы использовали известность дисперсии и то, что распределение случайной величины $n^{1/2}(\bar{x} - m)/\sigma$ есть $N(0, 1)$. Теперь мы будем считать, что дисперсия неизвестна, поэтому эту статистику при всем желании вычислить мы не можем. Естественно попытаться заменить ее на

$$T_m(x_1, x_2, \dots, x_n) = n^{1/2}(\bar{x} - m)/S,$$

где S^2 — несмешенная оценка дисперсии. И действительно, можно определить распределение статистики T_m , которое не зависит от неизвестной нам дисперсии и от m , и использовать T_m для построения доверительного интервала для среднего m . Имеем:

$$\begin{aligned} T_m &= n^{1/2} \frac{\bar{x} - m}{S} = \frac{(x_1 - m) + (x_2 - m) + \dots + (x_n - m)}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2}} = \\ &= \frac{1}{n^{1/2}} \frac{\left(\frac{x_1 - m}{\sigma} + \frac{x_2 - m}{\sigma} + \dots + \frac{x_n - m}{\sigma} \right)}{\left(\frac{1}{n} \sum_{i=1}^n \left[\frac{x_i - m}{\sigma} - \frac{\bar{x} - m}{\sigma} \right]^2 \right)^{1/2}}. \end{aligned}$$

Обозначим $\xi_i = \frac{x_i - m}{\sigma}$, тогда ξ_i независимы и имеют распределение $N(0, 1)$,

$$T_m = \frac{n^{1/2} (\xi_1 + \dots + \xi_n)}{\left(\frac{1}{n} \sum_{i=1}^n (\xi_i - \bar{\xi})^2 \right)^{1/2}}.$$

Как и при решении задачи 3 проведем вращение нормального случайного вектора (x_1, x_2, \dots, x_n) , мы уже видели, что

$$n^{1/2}(\xi_1 + \dots + \xi_n) = \eta_1, \frac{1}{n} \sum_{i=1}^n (\xi_i - \bar{\xi})^2 = \eta_2^2 + \dots + \eta_n^2.$$

Таким образом,

$$T_m = \frac{\eta_1}{\left(\frac{1}{n} \sum_{i=2}^n \eta_i^2 \right)^{1/2}}$$

имеет распределение Стьюдента с параметром $n - 1$, доверительный интервал определяется из неравенства $|T_m| < \varepsilon$, где $\mathbb{P}\{|T_m| < \varepsilon\} = \gamma$ а ε определяется по γ и таблице распределения Стьюдента.

Замечание 4. При больших n распределение Стьюдента хорошо аппроксимируется стандартным нормальным распределением.

Замечание 5. Мы видим, что в этой задаче, так же, как и в задачах 2 и 3, длина доверительного интервала случайна и зависит от результатов наблюдений. Это не очень соответствует высказанной выше идеи введения состоятельной оценки. В то же время оценка \bar{x} является состоятельной согласно закону больших чисел. Для эмпирической дисперсии выше была сформулирована задача о состоятельности, которую сейчас есть смысл решить.

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2 - (\bar{x} - m)^2.$$

Первое слагаемое сходится к σ^2 по вероятности в силу закона больших чисел, среднее второго слагаемого сходится к нулю, а в силу неотрицательности слагаемого оно сходится к нулю по вероятности.

Замечание 6. Из анализа примеров создается впечатление, что доверительный интервал с одним и тем же доверительным уровнем для всех значений параметра θ существует редко и лишь в особых ситуациях, в которых возникают инвариантные распределения, не зависящие от неизвестного нам значения параметра. Оказывается, это не так, и аналог доверительного интервала можно связать почти с любой статистикой T . Действительно, зафиксируем для любого θ два числа — $t_1(\theta)$ и $t_2(\theta)$ такие, что

$$\mathbb{P}_\theta^{(n)} \{t_1(\theta) < T < t_2(\theta)\} = \gamma. \quad (*)$$

(Гарантировать существование таких $t_1(\theta)$ и $t_2(\theta)$ можно лишь в том случае, когда функция распределения статистики T непрерывна при любом значении параметра θ .) Тогда при любом значении T в качестве аналога доверительного интервала мы можем взять множество всех θ таких, что выполняется (*). Действительно, при любом θ это значение параметра с вероятностью γ удовлетворяет условию (*) и попадает в выбранное множество. Разумеется, такое множество в конкретной модели может быть не интервалом (и даже не одномерным), поэтому множество таких θ корректнее называть доверительным множеством, но такого термина я не встречал.

В случае многомерного параметра θ приведенная конструкция имеет смысл лишь если мы оцениваем все координаты θ одновременно. Например, в нормальной модели при фиксированном m множество (*) будет разным при разных значениях σ .

3. Вывод функции плотности распределения Стьюдента. В теории вероятности мы отложили вывод распределения Стьюдента. Сейчас самое время этим заняться, хотя для пользователей математической статистики этот вывод не нужен, достаточно иметь таблицу функций распределения, а для составления таблиц нужно уметь хорошо приближать интегралы, которые заведомо не берутся в явном виде.

Итак, мы рассматриваем случайную величину

$$t_n = \frac{\xi}{\sqrt{\xi_1^2 + \dots + \xi_n^2}} = \frac{\xi}{\sqrt{\frac{1}{n}\chi_n^2}},$$

где ξ и $\xi_i - n+1$ независимых нормальных $N(0, 1)$ случайных величин. В частности, числитель и знаменатель случайной величины t_n независимы. При $n = 1$ величину t_n можно представить в виде $\frac{\xi}{|\eta|}$, где ξ и η независимы и нормальны $N(0, 1)$. Заметим, что если произвольную случайную величину ζ умножить на независимую двузначную случайную величину ε , которая задается соотношением $\mathbb{P}\{\varepsilon = 1\} = \mathbb{P}\{\varepsilon = -1\} = 1/2$, то получим новую случайную величину $\varepsilon\zeta$, распределение которой связано с распределением ζ соотношением

$$\mathbb{P}\{\varepsilon\zeta \in [-x, x]\} = 2\mathbb{P}\{\varepsilon\zeta \in [0, x]\} = \mathbb{P}\{\zeta \in [-x, x]\}.$$

Распределение со свойством $\mathbb{P}\{[-x, x]\} = 2\mathbb{P}\{[0, x]\}$ называется симметричным. Для симметричных распределений (например, для нормального $N(0, 1)$) таблицу значений функции распределения $F(x)$ достаточно задать лишь для $x > 0$. Если же случайная величина ζ сама имеет симметричное распределение, то легко видеть, что умножение на случайный знак ε не меняет распределение. Поэтому (объяснить!)

$$\frac{\xi}{|\eta|} \stackrel{L}{=} \frac{\varepsilon\xi}{|\eta|} \stackrel{L}{=} \frac{\xi}{\varepsilon|\eta|} \stackrel{L}{=} \frac{\xi}{\eta},$$

то есть имеет распределение Коши. Итак, при выводе распределения Стьюдента целесообразно использовать преобразование координат, которое использовалось при выводе распределения Коши. Но предварительно мы выпишем плотности распределения:

$$\begin{aligned} p_{\chi_n^2}(x) &= \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} e^{-\frac{x}{2}} x^{\frac{n}{2}-1} I_{(0,\infty)}(x), \\ p_{\frac{1}{n}\chi_n^2}(x) &= \frac{\left(\frac{n}{2}\right)^{n/2}}{\Gamma\left(\frac{n}{2}\right)} e^{-\frac{nx}{2}} x^{\frac{n}{2}-1} I_{(0,\infty)}(x), \\ p_{\sqrt{\frac{1}{n}\chi_n^2}}(x) &= 2\frac{\left(\frac{n}{2}\right)^{n/2}}{\Gamma\left(\frac{n}{2}\right)} e^{-\frac{nx^2}{2}} x^{n-1} I_{(0,\infty)}(x), \end{aligned}$$

Далее запишем функцию распределения t_n :

$$F_{t_n}(x) = \int_{u/v \leq x, v \geq 0} p_\xi(u)p_{\sqrt{\frac{1}{n}\chi_n^2}}(v)dudv = \int_{u/v \leq x, v \geq 0} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} 2\frac{\left(\frac{n}{2}\right)^{n/2}}{\Gamma\left(\frac{n}{2}\right)} e^{-\frac{nv^2}{2}} v^{n-1} dudv =$$

Перейдем к новым координатам $y = u/v$, $z = v$, тогда $\frac{\partial(u,v)}{\partial(x,y)} = z$:

$$= \int_{y \leq x, z \geq 0} \frac{1}{\sqrt{2\pi}} 2\frac{\left(\frac{n}{2}\right)^{n/2}}{\Gamma\left(\frac{n}{2}\right)} z^n e^{-\frac{1}{2}z^2(n+y^2)} dy dz =$$

$$(w = \frac{1}{2}z^2(n+y^2))$$

$$= \int_{y \leq x, w \geq 0} \frac{1}{\sqrt{\pi}} \frac{n^{n/2}}{\Gamma\left(\frac{n}{2}\right)} e^{-w} w^{\frac{n-1}{2}} (n+y^2)^{-\frac{n+1}{2}} dy dw = \int_{y \leq x} \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}} dy.$$

§20. Достаточные статистики.

При построении новых оценок обычно стараются использовать статистики, которые называются достаточными.

Определение. Статистика $T = T(x_1, \dots, x_n)$ называется *достаточной* для статистической модели $\mathbb{P}_\theta^{(n)}$, $\theta \in \Theta$, если (в непрерывной модели) условная плотность $p_\theta^{(n)}(x_1, \dots, x_n | T = t)$ или (в дискретной модели) условная вероятность $\mathbb{P}_\theta^{(n)}\{x_1, \dots, x_n | T = t\}$ не зависит от θ . Объяснить это определение можно следующим образом: если статистика достаточна, то в ней содержится вся информация выборки относительно параметра θ .

Замечание 1. Разумеется, статистика T может принимать и векторные значения.

Теорема. Если (в непрерывной модели) совместная плотность p_θ^n допускает факторизацию

$$p_\theta^{(n)}(x_1, \dots, x_n) = g_\theta(T(x_1, \dots, x_n))h(x_1, \dots, x_n),$$

или (в дискретной модели) совместная вероятность $\mathbb{P}_\theta^{(n)}\{x_1, \dots, x_n\}$ допускает факторизацию

$$\mathbb{P}_\theta^{(n)}\{x_1, \dots, x_n\} = g_\theta(T(x_1, \dots, x_n))h(x_1, \dots, x_n),$$

где функция $g_\theta(t)$ зависит от θ , а функция $h(x_1, \dots, x_n)$ не зависит от θ , то статистика T является достаточной.

Доказательство мы проведем лишь для дискретного случая, где оно наиболее просто. По определению условной вероятности имеем:

$$\mathbb{P}_\theta^{(n)}\{x_1, \dots, x_n | T = t\} = \frac{\mathbb{P}_\theta^{(n)}\{\{x_1, \dots, x_n\} \cap \{T = t\}\}}{\mathbb{P}_\theta^{(n)}\{T = t\}}.$$

В этой дроби числитель равен нулю, если $T(x_1, \dots, x_n) \neq t$, и равен

$$\mathbb{P}_\theta^{(n)}\{\{x_1, \dots, x_n\} = g_\theta(t)h(x_1, \dots, x_n),$$

если $T(x_1, \dots, x_n) = t$. Знаменатель считается суммированием:

$$\mathbb{P}_\theta^{(n)}\{T = t\} = \sum_{T(x_i)=t} \mathbb{P}_\theta^{(n)}\{x_1, \dots, x_n\} = \sum_{T(x_i)=t} g_\theta(t)h(x_1, \dots, x_n).$$

В результате дробь имеет вид

$$\frac{h(x_1, \dots, x_n)}{\sum_{T(x_i)=t} h(x_1, \dots, x_n)},$$

то есть не зависит от θ .

Замечание 2. В вероятностной модели с плотности (такое распределение мы называем непрерывным) возникает некоторая трудность при вычислении условной плотности, так как совместная плотность вектора (x_1, \dots, x_n) и статистики $T(x_1, \dots, x_n)$ не существует. Действительно, она должна была бы быть определена на $n + d$ -мерном пространстве, где d — размерность T , но вся вероятность на этом пространстве сосредоточена на множестве лебеговой меры нуль $\{(x_1, \dots, x_n, t) : t = T(x_1, \dots, x_n)\}$.

Пример 1. В нормальной модели

$$\begin{aligned} p_{m,\sigma}^{(n)}(x_1, x_2, \dots, x_n) &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^n \left(\frac{(x_i - m)}{\sigma}\right)^2\right) = \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2} \frac{\sum_{i=1}^n x_i^2 - 2m \sum_{i=1}^n x_i + nm^2}{\sigma^2}\right). \end{aligned}$$

Легко видеть, что условиям теоремы удовлетворяет статистика

$$T(x_1, \dots, x_n) = \left(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2 \right),$$

и что эмпирические среднее и дисперсию являются функциями этой статистики, в свою очередь, двумерная статистика T является функцией двумерной статистики (\bar{x}, s^2) .

Пример 2. Поучительный пример приведен в книге И.Н. Володина. Рассмотрим равномерное распределение на отрезке $[0, \Delta]$, где $\Delta > 0$ — неизвестный параметр. Совместная функция плотности задается формулой

$$p_{\Delta}^{(n)}(x_1, x_2, \dots, x_n) = \left(\frac{1}{\Delta} \right)^n I_{[0, \Delta]}(x_1) \cdots I_{[0, \Delta]}(x_n).$$

В этой модели достаточной статистикой является статистика

$$T(x_1, \dots, x_n) = \max_{i \leq n} x_i.$$

Действительно, если $T \leq \Delta$, то

$$p_{\Delta}^{(n)}(x_1, x_2, \dots, x_n) = \left(\frac{1}{\Delta} \right)^n,$$

если $T > \Delta$, то

$$p_{\Delta}^{(n)}(x_1, x_2, \dots, x_n) = 0.$$

Все же оба этих примера являются примерами 'второго сорта', в них функцию h можно считать равной константе, например, 1. Иная ситуация в модели Пуассона:

Пример 3.

$$P_{\lambda}^{(n)}(k_1, k_2, \dots, k_n) = \frac{\lambda^{k_1+k_2+\dots+k_n}}{k_1!k_2!\cdots k_n!} e^{-n\lambda}.$$

В этом примере $T(k_1, k_2, \dots, k_n) = k_1 + k_2 + \dots + k_n$,

$$h(k_1, k_2, \dots, k_n) = \frac{1}{k_1!k_2!\cdots k_n!}.$$

§21. Сравнение двух гипотез

Рассматриваются две гипотезы: гипотеза H_0 — 'истинное распределение — \mathbb{P}_0 с функцией плотности p_0 ', и гипотеза H_1 — 'истинное распределение — \mathbb{P}_1 с функцией плотности p_1 '. На основании выборки (x_1, \dots, x_n) (x_i имеют распределение \mathbb{P}_0 или \mathbb{P}_1) нам следует выбрать одну из гипотез — H_0 или H_1 .

Естественно, что критерий различия двух гипотез имеет вид $\mathbf{R}^n = \Delta_0 + \Delta_1$ (\mathbf{R}^n в данном случае — пространство всех выборок), гипотеза H_0 принимается, если выборка окажется в Δ_0 , H_1 принимается, если выборка окажется в Δ_1 . Нас интересует такое разбиение $\Delta_0 + \Delta_1$, при котором минимальны вероятности ошибок. Отметим, что

$\mathbb{P}\{\text{принять гипотезу } H_1 \text{ при условии, что верна гипотеза } H_0\} = \alpha$ называется вероятностью ошибки первого рода, $\alpha = \mathbb{P}_0^{(n)} \Delta_1$,

$\mathbb{P}\{\text{принять гипотезу } H_0 \text{ при условии, что верна гипотеза } H_1\} = \beta$ называется вероятностью ошибки второго рода, $\beta = \mathbb{P}_1^{(n)} \Delta_0$.

Эта терминология придает разный смысл гипотезам H_0 и H_1 , гипотеза H_0 считается основной, а H_1 — альтернативной, поэтому α — вероятность отвергнуть правильную гипотезу (т.е.

H_0), β — вероятность принять неправильную гипотезу (тоже H_0). Оптимальность выбора разбиения $\Delta_0 + \Delta_1$ обычно означает фиксацию вероятности α и минимизацию при фиксированном α вероятности β .

Лемма Неймана — Пирсона. *Если*

$$\mathbb{P}_0 \left\{ (x_i) : \frac{p_0^{(n)}(x_1, \dots, x_n)}{p_1^{(n)}(x_1, \dots, x_n)} = C \right\} = 0, \quad (0)$$

где константа C определяется из равенства

$$\int_{\Delta_1} \cdots \int p_0^{(n)}(x_1, \dots, x_n) dx_1 \dots dx_n = \alpha = \mathbb{P}\{H_1 | H_0\},$$

то оптимальный критерий имеет вид

$$\Delta_0 = \left\{ (x_i) : \frac{p_0^{(n)}(x_1, \dots, x_n)}{p_1^{(n)}(x_1, \dots, x_n)} > C \right\}, \Delta_1 = \mathbf{R}^n \setminus \Delta_0. \quad (*)$$

Доказательство. Сравним с критерием (*) любой другой критерий $\mathbf{R}^n = \Delta'_0 + \Delta'_1$, где также $\int_{\Delta'_1} \cdots \int p_0^{(n)}(x_1, \dots, x_n) dx_1 \dots dx_n = \alpha$. Имеем:

$$1 - \alpha = \mathbb{P}_0^{(n)} \Delta_0 = \mathbb{P}_0^{(n)}(\Delta_0 \cap \Delta'_0) + \mathbb{P}_0^{(n)}(\Delta_0 \setminus \Delta'_0) = \mathbb{P}_0^{(n)} \Delta'_0 = \mathbb{P}_0^{(n)}(\Delta_0 \cap \Delta'_0) + \mathbb{P}_0^{(n)}(\Delta'_0 \setminus \Delta_0).$$

Поэтому

$$\mathbb{P}_0^{(n)}(\Delta_0 \setminus \Delta'_0) = \mathbb{P}_0^{(n)}(\Delta'_0 \setminus \Delta_0).$$

Мы используем следующие очевидные равенства:

$$\Delta_0 \setminus \Delta'_0 = \Delta'_1 \setminus \Delta_1,$$

$$\Delta'_0 \setminus \Delta_0 = \Delta_1 \setminus \Delta'_1.$$

В итоге имеем:

$$\begin{aligned} \mathbb{P}_1^{(n)} \Delta_0 &= \mathbb{P}_1^{(n)}(\Delta_0 \cap \Delta'_0) + \mathbb{P}_1^{(n)}(\Delta_0 \setminus \Delta'_0) = \\ &= \mathbb{P}_1^{(n)}(\Delta_0 \cap \Delta'_0) + \int_{\Delta_0 \setminus \Delta'_0} \cdots \int p_1^{(n)} dx_1 \dots dx_n = \\ &= \mathbb{P}_1^{(n)}(\Delta_0 \cap \Delta'_0) + \int_{\Delta_0 \setminus \Delta'_0} \cdots \int \frac{p_1^{(n)}}{p_0^{(n)}} p_0^{(n)} dx_1 \dots dx_n \leq \\ &\leq \mathbb{P}_1^{(n)}(\Delta_0 \cap \Delta'_0) + \int_{\Delta_0 \setminus \Delta'_0} \cdots \int \frac{1}{C} p_0^{(n)} dx_1 \dots dx_n = \\ &= \mathbb{P}_1^{(n)}(\Delta_0 \cap \Delta'_0) + \int_{\Delta'_0 \setminus \Delta_0} \cdots \int \frac{1}{C} p_0^{(n)} dx_1 \dots dx_n \leq \\ &\leq \mathbb{P}_1^{(n)}(\Delta_0 \cap \Delta'_0) + \int_{\Delta'_0 \setminus \Delta_0} \cdots \int \frac{p_1^{(n)}}{p_0^{(n)}} p_0^{(n)} dx_1 \dots dx_n = \mathbb{P}_1^{(n)} \Delta'_0, \end{aligned}$$

что и требовалось.

Эта простая лемма (если подумать, то она очевидна) используется при исследовании критериев в задаче проверки гипотезы (см. следующую лекцию).

В ситуации, когда вероятность в (0) не равна 0, вероятность ошибки первого рода может перескочить через заданное нами α . Поэтому задачу решает рандомизированный критерий Неймана — Пирсона. Суть критерия состоит в том, что если имеет место событие в равенстве (0), мы 'подбрасываем монету' и в зависимости от результата эксперимента выбираем либо H_0 либо H_1 . Монета 'изготавливается' так, что при этом вероятность ошибки первого рода будет в точности равна α .

Термин 'рандомизация' в переводе с английского означает ослучайнивание.

Переставить лекции??

§22. Задача проверки гипотез.

В этой лекции будет идти речь о проверки вероятностных гипотез. Имеется в виду, что по результатам наблюдений мы должны вынести суждение (для заказчика) — согласуются результаты наблюдений с выдвигаемой им (или вами) гипотезой (или нет). Вначале для простоты мы будем считать, что гипотеза заключается в задании распределения наблюдаемой случайной величины. Итак, наблюдаются числа x_1, \dots, x_n , мы проверяем гипотезу о том, что это — результаты наблюдений независимых случайных величин с одним и тем же распределением \mathbb{P}_0 . Критерий имеет вид $\Delta \subset \mathbf{R}^n$, это множество называется *критическим*, гипотеза отвергается, если мы (т.е. (x_1, \dots, x_n)) попадем в критическое множество (точнее, мы говорим, что в этом случае результаты наблюдений не согласуются с выдвинутой гипотезой). Само критическое множество выбирается по *уровню значимости* α — вероятности попадания в критическое множество при условии справедливости гипотезы (обычно α равно одному из чисел 0.001, 0.01, 0.05, 0.1), именно для этих значений α обычно составляются вероятностные таблицы. Итак, при условии справедливости \mathbb{P}_0 вероятность попадания в критическое множество мала, при проверке гипотез мы исходим из убеждения, что события с малой вероятностью как правило не происходят, если же такое событие произошло, то, скорее всего, наше первоначальное предположение (о малости вероятности попадания в Δ) неверно, а следовательно, неверна и гипотеза, при условии справедливости которой была сосчитана вероятность попадания в Δ .

Замечание. $\alpha = \mathbb{P}_0^{(n)}(\Delta)$.

1. Примеры.

Пример (который должен заставить задуматься!). Пусть мы подбрасываем большое число раз монету (например, 10000 раз), гипотеза состоит в том, что вероятность выпадения герба в точности равна половине, т.е. монета симметрична. Как можно выбрать критическое множество (для простоты вычислений уровень значимости будет считаться равным 1/125 примерно). Обозначим ν — частота выпадения герба. 1 вариант критического множества — $\Delta = \{\nu > 1 - \varepsilon\}$, где ε выбирается так, что

$$\mathbb{P}_0^{(n)}\{\nu > 1 - \varepsilon\} = 1/125.$$

Второй вариант — $\tilde{\Delta} = \{\nu < \varepsilon\}$,

$$\mathbb{P}_0^{(n)}\tilde{\Delta} = 1/125.$$

Можно также придумать симметричный критерий: $\Delta^\# = \{|\nu - 1/2| > \delta\}$, где δ подбирается так, что $\mathbb{P}_0^{(n)}\Delta^\# \approx 1/125$.

Но оказывается возможен вообще парадоксальный критерий $\Delta' = \{\nu = 1/2\}$, по формуле Стирлинга $\mathbb{P}_0^{(n)}\Delta' \approx 1/125$.

Критерий Δ' не противоречит формальному определению критерия, вероятность попадания в Δ' мала, в то же время получается, что мы отвергаем гипотезу, если она слишком хорошо подтверждается: $\nu = 1/2$. Итак критериев много, нужно не только уметь их придумывать, но и уметь сравнивать, а для этого кроме основной гипотезы \mathbb{P}_0 нужен также набор альтернативных гипотез (в данном случае это гипотезы \mathbb{P}_p , где $\mathbb{P}_p\{0\} = p, \mathbb{P}_p\{1\} = 1 - p$). Обозначим:

$$\beta(p) = \mathbb{P}_p\Delta'^c,$$

$$\begin{aligned}\tilde{\beta}(p) &= \mathbb{P}_p \tilde{\Delta}^c, \\ \beta^\#(p) &= \mathbb{P}_p \Delta^{\#c}, \\ \beta'(p) &= \mathbb{P}_p \Delta'^c.\end{aligned}$$

Мы будем считать, что один критерий лучше другого, если функция $\beta(p)$ для одного критерия всегда больше чем аналогичная функция для другого критерия. При таком способе сравнения выбор оптимального критерия зависит от выбора набора альтернатив $\{\mathbb{P}_p\}$. Если этот набор совпадает с $\{\mathbb{P}_p : p > 1/2\}$, то, нетрудно видеть, что оптимальным является критерий Δ , если $\{\mathbb{P}_p : p < 1/2\}$, то оптимален критерий $\tilde{\Delta}$, если $\{\mathbb{P}_p : p \neq 1/2\}$, то вообще не будет оптимального критерия (это легко проверить). Ни для одной из этих альтернатив критерий Δ' не будет хорошим (это тоже легко проверяется).

Пример. 1000 раз подбрасывали монету, 550 раз выпал герб. Согласуется ли этот результат с гипотезой $p = 1/2$? Критическое множество разумно выбрать в виде $\Delta = \{k \geq 550\}$, где k — число гербов в 1000 испытаниях, тогда

$$\alpha = \mathbb{P}(\Delta) = \sum_{550}^{1000} C_{1000}^k p^k (1-p)^{1000-k} = \sum_{550}^{1000} C_{1000}^k \left(\frac{1}{2}\right)^{1000}.$$

Эту вероятность можно явно вычислить на компьютере, но при этом надо избегать машинных нулей и бесконечностей. Можно воспользоваться и нормальным приближением. Согласно центральной предельной теореме, распределение случайной величины k примерно нормальному, нужно свести вопрос к таблицам нормального распределения $N(0, 1)$ — функции Φ . Для этого надо вычесть из k среднее и разделить разность на корень квадратный из дисперсии. Имеем

$$\mathbb{E}k = 1000 \cdot \frac{1}{2} = 500, \mathbb{D}k = 1000 \cdot \frac{1}{4} = 250.$$

Итак,

$$\mathbb{P}\{k \geq 550\} \approx \Phi(50/\sqrt{250}) \approx \Phi(3.164) \leq 0.001.$$

Гипотеза не подтвердилась.

Задача выбора критического множества существенно усложняется, если наша гипотеза является *параметрической*. Нужно стремиться к тому, что вероятность попадания в критическое множество была бы инвариантна — не зависела бы от неизвестных нам значений параметра.

Замечание. Чаще всего критическое множество имеет вид

$$\Delta = \{T > t\},$$

где T — некоторая статистика (в параметрическом случае ее распределение не должно зависеть от параметра). В такой ситуации процедура проверки приобретает другой вид — мы вычисляем конкретное значение T и узнаем (обычно по таблице) уровень значимости, для которого наша гипотеза не подтвердилась. В результате мы можем получить любое число между 0 и 1, которое не обязательно равно 0.01, 0.05 или 0.1. Такой подход иногда удобнее. Результаты вычислений становятся более информативными, кроме того, вместо таблиц мы можем иногда использовать компьютерные вычисления (см. Ранговые критерии). Действительно, мы вычислили значение T для нашей конкретной серии экспериментов, получили некоторое значение \tilde{T} , и теперь нас не интересует вся таблица значений функции распределения статистики T , а лишь конкретное число $\mathbb{P}\{T > \tilde{T}\}$.

2. Критерий Стьюдента.

Даны две выборки — (x_1, \dots, x_n) и (y_1, \dots, y_k) . Мы предполагаем, что одна выборка получена в ходе одних экспериментов, а другая — в ходе других экспериментов, но при этом измерялось одно и то же. В результате мы получили, что среднее одной выборки, \bar{x} больше среднего другой выборки \bar{y} , $\bar{x} > \bar{y}$. Возникает естественный вопрос, случайно это превышение или нет,

не означает ли оно, что на самом деле в двух сериях экспериментов мы измеряли разные характеристики.

Пример из истории химии.

Двумя способами измерялся удельный вес воздуха. В первой серии экспериментов воздух фиксированного объема воздух сжижался и после этого взвешивался, были получены числа (x_1, \dots, x_n) . Во второй серии отдельные химические вещества в воздухе измерялись благодаря участию в химических реакциях и их вес складывался. Таким образом были получены числа (y_1, \dots, y_k) . В итоге выяснилось, что $\bar{x} - \bar{y} = \varepsilon > 0$. Не означает ли это, что в воздухе присутствуют вещества, не вступающие в химические реакции? В дальнейшем такие вещества были открыты, это — инертные газы. Итак, возникает вопрос: почему равна вероятность $\mathbb{P}\{\bar{x} - \bar{y} \geq \varepsilon\}$, где ε вычислена раньше из двух конкретных выборок, полученных в результате конкретного эксперимента, а \bar{x} и \bar{y} — средние произвольных случайных выборок? Если эта вероятность мала, то мы можем думать, что превышение \bar{x} над \bar{y} не случайно, и на самом деле среднее

$$m_x = \mathbb{E}x_i > m_y = \mathbb{E}y_i.$$

Но в вероятностной модели, в рамках которой мы считаем $\mathbb{P}\{\bar{x} - \bar{y} > \varepsilon\}$, мы предполагаем, что $m = m_x = m_y$, $\sigma^2 = \mathbb{D}x_i = \mathbb{D}y_i$, случайные величины x_i, y_j — независимые нормальные случайные величины с одними и теми же параметрами m, σ . Если в рамках такой модели вероятность $\mathbb{P}\{\bar{x} - \bar{y} > \varepsilon\}$ будет мала, то будет естественно предположить, что модель не верна, а на самом деле $m_x > m_y$. Все это проделал химик Стьюдент (псевдоним Госсета). К сожалению, у меня нет данных Стьюдента, но по-видимому, не все x_i были больше всех y_j . Иначе утверждение $m_x > m_y$ стало бы очевидным и без использования критерия Стьюдента. Как и при построении доверительного интервала для среднего при неизвестной дисперсии, число ε будет случайным. Точнее, малое событие имеет вид

$$\left\{ C(n, k) \frac{\bar{x} - \bar{y}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{j=1}^k (y_j - \bar{y})^2}} > \delta \right\}.$$

Оказалось, что при некотором выборе константы $C(m, k)$ случайная величина

$$C(n, k) \frac{\bar{x} - \bar{y}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{j=1}^k (y_j - \bar{y})^2}}$$

имеет распределение Стьюдента с параметром $n + k - 2$, то есть распределение случайной величины

$$t_{n+k-2} = \frac{\xi}{\sqrt{\frac{1}{n+k-2} \sum_{i=1}^{n+k-2} \xi_i^2}},$$

в частности, распределение этой величины не зависит от параметров m и σ , а число δ может быть найдено по таблице распределения случайной величины t_{n+k-2} . Для доказательства используются простые выкладки, к которым мы уже привыкли.

$$\begin{aligned} C(n, k) \frac{\bar{x} - \bar{y}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{j=1}^k (y_j - \bar{y})^2}} &= \\ C(n, k) \frac{\frac{\bar{x} - m}{\sigma} - \frac{\bar{y} - m}{\sigma}}{\sqrt{\sum_{i=1}^n \left(\frac{[x_i - m] - [\bar{x} - m]}{\sigma} \right)^2 + \sum_{j=1}^k \left(\frac{[y_j - m] - [\bar{y} - m]}{\sigma} \right)^2}} &= \end{aligned}$$

далее мы введем независимые случайные величины с распределением $N(0, 1)$:

$$\xi_i = \frac{x_i - m}{\sigma}, \eta_j = \frac{y_j - m}{\sigma},$$

также

$$\bar{\xi} = \frac{\xi_1 + \dots + \xi_n}{n}, \bar{\eta} = \frac{\eta_1 + \dots + \eta_k}{k}.$$

Теперь мы продолжим наше равенство:

$$\begin{aligned} &= C(n, k) \frac{\bar{\xi} - \bar{\eta}}{\sqrt{\sum_{i=1}^n [\xi_i - \bar{\xi}]^2 + \sum_{j=1}^k [\eta_j - \bar{\eta}]^2}} = \\ &= C(n, k) \frac{\bar{\xi} - \bar{\eta}}{\sqrt{\sum_{i=1}^n \xi_i^2 - n\bar{\xi}^2 + \sum_{j=1}^k \eta_j^2 - k\bar{\eta}^2}}. \end{aligned}$$

Далее мы используем вращение, после которого первая координата имеет вид $\frac{x_1 + \dots + x_n}{\sqrt{n}}$, а вторая координата имеет вид $\frac{y_1 + \dots + y_k}{\sqrt{k}}$, при этом случайный вектор $(\xi_1, \dots, \xi_n, \eta_1, \dots, \eta_k)$ превратится в вектор

$$\left(\zeta_1 = \frac{\xi_1 + \dots + \xi_n}{\sqrt{n}}, \zeta_2 = \frac{\eta_1 + \dots + \eta_k}{\sqrt{k}}, \zeta_3, \dots, \zeta_{n+k} \right).$$

Числитель и знаменатель оказываются независимыми между собой. Числитель равен $\frac{\zeta_1}{\sqrt{n}} - \frac{\zeta_2}{\sqrt{k}}$, а знаменатель $= \sqrt{\sum_{i=1}^{n+k} \zeta_i^2 - \zeta_1^2 - \zeta_2^2}$. Знаменатель дает нам в $C(n, k)$ множитель $\sqrt{n+k-2}$, а числитель имеет дисперсию $1/n + 1/k$, поэтому дает множитель $\sqrt{\frac{nk}{n+k}}$.

Замечание. Критерий Стьюдента вошел в парадигму многих гуманитарных и естественных дисциплин. Вопросы 'Что дала проверка по Стьюденту?', 'А Вы проверили эффективность Вашего метода по Стьюденту?' часто задаются на защитах докторских и кандидатских диссертаций по медицине, психологии, социологии и даже педагогике. Положительный ответ вызывает одобрение. При этом имеется в виду сравнение результатов, полученных для одной группы новым, предлагаемым докторантам методом, со старым, испытаным методом для контрольной группы. Попадание в критическую область интерпретируется как проверка эффективности нового метода. Во многих случаях применение нормальной модели не только не обосновано, но и заведомо неверно, поэтому при подсчете вероятности попадания в критическое множество можно ошибиться в несколько раз. Для малых выборок мы не вправе мыслить и в рамках центральной предельной теоремы и вытекающей из нее нормальной модели. Но если уровень значимости критерия близок к нулю, то это не страшно. Действительно, пусть даже произошло событие с вероятностью не 0.01, а например 0.03. Все равно эта вероятность мала и результаты проверки свидетельствуют о преимуществах новой методики. Математик находится в таких случаях в худшем положении. Он прекрасно понимает, что использование распределения Стьюдента некорректно, но ничего лучшего у него все равно нет.

Из сделанных выкладок следует знаменитая

Теорема Фишера. В нормальной модели эмпирическое среднее \bar{x} и эмпирическая дисперсия s^2 — независимые случайные величины.

2. Критерий χ^2 .

Критерий χ^2 проверяет гипотезу

$$\mathbb{P}(A_1) = p_1, \dots, \mathbb{P}(A_r) = p_r \text{ где } \Omega = A_1 + \dots + A_r.$$

Пусть мы провели n экспериментов, в n_i экспериментах произошло событие A_i ($n_1 + \dots + n_r = n$). Из многомерной центральной предельной теоремы вытекает, что при больших n случайная

величина

$$T = \sum_{j=1}^r \left(\frac{n_j - np_j}{\sqrt{np_j}} \right)^2$$

имеет распределение, близкое распределению случайной величины χ_{r-1}^2 . Критерий строится в виде

$$\Delta = \left\{ T = \sum_{j=1}^r \left(\frac{n_j - np_j}{\sqrt{np_j}} \right)^2 > \varepsilon \right\}, \quad (*)$$

где число ε определяется по уровню значимости по таблицам распределения χ_{r-1}^2 .

Важное замечание. Критерий χ^2 применим и в ситуации, когда сами вероятности p_i являются функциями s параметров $\alpha_1, \dots, \alpha_s$. В этом случае надо предварительно оценить параметры α_i так, чтобы сумма T оказалась бы минимальной (в сложных ситуациях это можно сделать только приближенно). Доказывается (Крамер, гл. 30, п. 3), что при подстановке этих оценок в выражение для T случайная величина T будет иметь распределение χ_{r-s-1}^2 . Заодно мы сообщили также еще об одном универсальном методе оценивания параметров — *методе минимума χ^2* . В книге Крамера написано также об *упрощенной процедуре вычисления минимума χ^2* . Дело в том, что при дифференцировании системы (*) по параметрам может получиться система уравнений, которую очень трудно разрешить. Предлагается упрощенная процедура, в рамках которой мы дифференцируем в (*) лишь числители. Оказывается (см. Крамер), это приводит при больших n к небольшой ошибке. Там же доказывается теорема о существовании и единственности при выполнении определенных условий решения возникающей системы и о сходимости распределения вычисляемой после подстановки в (*) этих решений случайной величины к распределению χ_{r-s-1}^2 . Пример применения критерия χ^2 будет дан ниже, в разделе проверки на независимость.

Мы приведем обоснование критерия χ^2 только в непараметрической ситуации. Сначала рассмотрим случайный вектор

$$(\eta_j) = \left(\frac{n_j - np_j}{\sqrt{np_j}} \right).$$

Очевидно, что

$$\mathbb{E} \left(\frac{n_j - np_j}{\sqrt{np_j}} \right) = 0$$

для всех j . Введем n независимых при различных $i \leq n$ двузначных случайных величин

$$\delta_j^{(i)} = \begin{cases} 1 & \text{с вероятностью } p_j \\ 0 & \text{с вероятностью } 1 - p_j \end{cases}$$

таких, что

$$\delta_j^{(i)} \delta_k^{(i)} = 0 \text{ при } j \neq k.$$

Тогда случайные величины n_j можно представить в виде

$$n_j = \sum_i \delta_j^{(i)},$$

а случайные величины η_j как

$$\eta_j = \frac{\sum_{i=1}^n \frac{\delta_j^{(i)} - p_j}{\sqrt{p_j}}}{\sqrt{n}}.$$

Итак, случайный вектор (η_j) можно представить как деленную на \sqrt{n} сумму n независимых случайных векторов $\frac{\delta_j^{(i)} - p_j}{\sqrt{p_j}}$ с нулевым средним. При большом числе экспериментов мы можем считать вектор (η_j) имеющим совместное нормальное распределение. Действительно, согласно многомерной центральной предельной теореме, распределение суммы сходится к нормальному распределению с той же матрицей вторых центральных смешанных моментов, что и у вектора $\frac{\delta_j^{(i)} - p_j}{\sqrt{p_j}}$, где i — некоторый номер эксперимента, например, $i = 1$. Вычислим эту матрицу, которую мы обозначим $[\mu_{jk}]_{j,k=1}^n$. При $k = j$ случайная величина $\delta_j^{(1)} \delta_k^{(1)}$ тождественно равна 1, поэтому

$$\mu_{jj} = \mathbb{E} \frac{(\delta_j^{(1)})^2 - p_j^2}{p_j} = \frac{p_j - p_j^2}{p_j} = 1 - p_j.$$

При $k \neq j$ случайная величина $\delta_j^{(1)} \delta_k^{(1)}$ тождественно равна 0, поэтому

$$\mu_{jk} = \mathbb{E} \frac{\delta_j^{(1)} \delta_k^{(1)} - p_j p_k}{\sqrt{p_j p_k}} = -\sqrt{p_j p_k}.$$

Резюмируем: (η_j) — случайные величины с нулевым средним, с выписанной выше матрицей вторых центральных смешанных моментов и с совместным нормальным распределением. Нас интересует распределение случайной величины $\sum_{j=1}^r \eta_j^2$. Чтобы найти это распределение, не занимаясь слишком много линейной алгеброй, мы используем искусственный прием. А именно, мы рассмотрим вектор (ξ_j) независимых нормальных случайных величин $N(0, 1)$, т.е. с единичной матрицей вторых центральных смешанных моментов и небольшим изменением этого вектора получим другой вектор с той же матрицей вторых центральных смешанных моментов, что и вектор (η_j) , поэтому мы будем иметь право считать этот вектор вектором (η_j) .

Для любого j положим

$$\eta_j = \xi_j - \sqrt{p_j} (\sqrt{p_1} \xi_1 + \sqrt{p_2} \xi_2 + \dots + \sqrt{p_r} \xi_r).$$

Заметим, что $\mathbb{E} (\sqrt{p_1} \xi_1 + \dots + \sqrt{p_r} \xi_r)^2 = p_1 + \dots + p_r = 1$. Имеем:

$$\begin{aligned} \mathbb{E} \eta_j^2 &= \mathbb{E} \xi_j^2 - 2\mathbb{E} \xi_j \sqrt{p_j} (\sqrt{p_1} \xi_1 + \dots + \sqrt{p_r} \xi_r) + \mathbb{E} p_j (\sqrt{p_1} \xi_1 + \dots + \sqrt{p_r} \xi_r)^2 = \\ &= 1 - 2p_j + p_j = 1 - p_j. \\ \mathbb{E} \eta_j \eta_k &= \mathbb{E} \xi_j \xi_k - \mathbb{E} \xi_j \sqrt{p_k} (\sqrt{p_1} \xi_1 + \dots + \sqrt{p_r} \xi_r) - \\ &\quad - \mathbb{E} \xi_k \sqrt{p_j} (\sqrt{p_1} \xi_1 + \dots + \sqrt{p_r} \xi_r) + \mathbb{E} \sqrt{p_j p_k} (\sqrt{p_1} \xi_1 + \dots + \sqrt{p_r} \xi_r)^2 = \\ &= 0 - \sqrt{p_k p_j} - \sqrt{p_j p_k} + \sqrt{p_j p_k} = -\sqrt{p_j p_k}. \end{aligned}$$

Далее мы рассмотрим сумму $\sum_{j=1}^r \eta_j^2$, преобразуем ее с помощью ортогонального преобразования $(\zeta_n) = U(\xi_n)$, где по лемме о вращении стандартного нормального вектора ζ_n независимы и нормальны $N(0, 1)$, а

$$\zeta_1 = \sqrt{p_1} \xi_1 + \dots + \sqrt{p_r} \xi_r.$$

Итак,

$$\begin{aligned} \sum_{j=1}^r \eta_j^2 &= \\ &= \sum_{j=1}^r \xi_j^2 - 2 \sum_{j=1}^r \xi_j \sqrt{p_j} (\sqrt{p_1} \xi_1 + \sqrt{p_2} \xi_2 + \dots + \sqrt{p_r} \xi_r) + \end{aligned}$$

$$\begin{aligned}
& + (\sqrt{p_1}\xi_1 + \dots + \sqrt{p_r}\xi_r)^2 = \sum_{j=1}^r \xi_j^2 - (\sqrt{p_1}\xi_1 + \dots + \sqrt{p_r}\xi_r)^2 = \\
& = \sum_{j=1}^r \zeta_j^2 - \zeta_1^2 = \sum_{j=2}^r \zeta_j^2.
\end{aligned}$$

§23. Обзор статистических критериев

Мы в основном приводим статистики, на которых основаны те или иные критерии. С выводом распределений или предельных распределений этих статистик можно познакомиться в специальной литературе.

1. Критерий Колмогорова — Смирнова.

Проверяет гипотезу о том, что выборка $\{x_1, \dots, x_n\}$ принадлежит генеральной совокупности с распределением F , где функция распределения F непрерывна. Основан на статистике

$$T = \sqrt{n} \sup_x |F_\Theta(x) - F(x)|,$$

где F_Θ — эмпирическая функция распределения. Критерий обычно имеет вид $\{T > \lambda\}$, где λ вычисляется по таблицам. Впрочем, для большого числа наблюдений известна асимптотическая формула Колмогорова: при $x > 0$

$$\mathbb{P}\{T < x\} \approx K(x) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 x^2}.$$

Впрочем, предпочтительнее использовать таблицы. Заметим, что распределение статистики T не зависит от функции F . Легко показывается, что мы можем принять в качестве $F(x) = x$ ($0 < x < 1$), то есть функцию равномерного распределения на $[0, 1]$. Дело в том, что

$$\mathbb{P}\{F(x_i) < x\} = \mathbb{P}\{x_i < F^{-1}(x)\} = F(F^{-1}(x)) = x.$$

Таким образом, преобразование аргумента $x \rightarrow F(x)$ приводит к равномерному распределению величин x_i (которые переходят в $F(x_i)$). Очевидно, что это преобразование не меняет величину T .

Можно в качестве популярного текста рассказать о выводе формулы Колмогорова. В следующем семестре мы будем изучать теорию случайных процессов и найдем распределение максимума случайного блуждания. Точно так же, с помощью метода отражения, находится распределение максимума процесса одномерного броуновского движения. Но нас интересует распределение максимума абсолютной величины, то есть вероятность захода движения вверх за уровень x или вниз за уровень $-x$, то есть вероятность объединения двух событий, при вычислении которой мы должны знать и вероятность пересечения этих событий. При этом приходится иметь дело с многократными пересечениями обоих уровней, и это приводит к ряду.

Имеется и другая тонкость: зависящий от времени $x \in [0, 1]$ процесс $\sqrt{n} \{F_\Theta(x) - F(x)\}$ (после преобразования времени) сходится не к броуновскому процессу $w(x)$, а к так называемому броуновскому мосту $w(x) - xw(1)$, который в момент времени $x = 1$ должен вернуться в 0. Действительно, в случае равномерного распределения на $[0, 1]$, к которому все сводится, $F_\Theta(1) - F(1) = 0$.

2. ω^2 -критерий Крамера-Мизеса.

3. Проверка однородности

Пусть имеются две выборки $\{x_1, x_2, \dots, x_n\}, \{y_1, y_2, \dots, y_k\}$. Задача состоит в проверке однородности объединения двух выборок, т.е. проверяется гипотеза о том, что обе выборки выбраны из одной генеральной совокупности.

Мы обсудим несколько возможных критериев проверки этой гипотезы.

Критерий Смирнова.

Основан на статистике

$$T = \sqrt{n} \sup_u |F_{\Theta,X}(u) - F_{\Theta,Y}(u)|,$$

где $F_{\Theta,X}$, $F_{\Theta,Y}$ — эмпирические функции распределения, построенные соответственно по результатам наблюдений $\{x_1, x_2, \dots, x_n\}$ и $\{y_1, y_2, \dots, y_k\}$. Критерий обычно имеет вид $\{|T| > \lambda\}$, где λ вычисляется по таблицам.

Критерий χ^2

Основан на статистике

$$T = \sum_{j=1}^r \frac{(n_j - np_j)^2}{np_j} + \sum_{j=1}^r \frac{(k_j - kp_j)^2}{kp_j}. \quad (1)$$

Здесь мы разбиваем множество исходов на r событий A_j , n_j — число попаданий x_i в событие A_j , k_j — число попаданий y_i в событие A_j , p_j — неизвестные нам вероятности A_j , причем $p_1 + \dots + p_r = 1$. Проверяется гипотеза, что в обеих выборках мы наблюдали случайные величины с одними и теми же вероятностями попадания в A_j . Неизвестные параметры p_j (их $r-1$) находятся по методу минимума χ^2 , (1) нужно продифференцировать по неизвестным параметрам и производные приравнять нулю. Причем используется упрощенная процедура вычисления минимума (т.е. дифференцируются по параметрам лишь числители в сумме). Проверьте сами, что решением системы уравнений является следующая сумма

$$\chi^2 = nk \sum_{j=1}^r \frac{1}{n_j + k_j} \left(\frac{n_j}{n} - \frac{k_j}{k} \right)^2.$$

Заметим, что число параметров равно $r-1$, точки минимума проще не вычислять, а угадать, они равны $p_i = \frac{n_i + k_i}{n+k}$. Согласно общему правилу предельное распределение имеет вид χ_s^2 , где $2r-2-(r-1)=r-1$. (Как и выше, из $2r$ вычитается 2 ввиду зависимостей возводимых в квадрат слагаемых: $\sum_i n_i = n$, $\sum_i k_i = k$.) Итак, критерий имеет вид $\{\chi^2 > \lambda\}$, где λ вычисляется из таблицы распределения χ_{r-1}^2 .

Критерий серий

Объединим выборки $\{x_1, x_2, \dots, x_n\}$, $\{y_1, y_2, \dots, y_k\}$ в одну и построим вариационный ряд. Напомним, что вариационным рядом $\{x_1^*, \dots, x_n^*\}$ для выборки $\{x_1, \dots, x_n\}$ называется сама выборка, но расположенная в порядке возрастания. Вариационный ряд для объединения состоит из $n+k$ чисел, расположенных в порядке возрастания. Запишем этот ряд как набор из $n+k$ символов x и y . Любой максимальный поднабор из расположенных рядом символов x мы будем называть серией. Критерий серий основан на статистике — числе таких серий. В справочниках можно найти инструкцию по вычислению функции распределения числа серий в предположении, что x и y имеют одно и то же распределение.

Критерий Вилкоксона

Рангом x_i называется номер наблюдения x_i в вариационном ряде $\{x_1^*, x_2^*, \dots, x_n^*\}$. Снова рассмотрим вариационный ряд объединения выборок $\{x_1, x_2, \dots, x_n\}$ и $\{y_1, y_2, \dots, y_k\}$ в одну и обозначим через X_i — ранг x_i в этом ряду. Критерий Вилкоксона основан на статистике

$$T = X_1 + \dots + X_n$$

и имеет вид $\left\{ \left| T - \frac{n(n+k+1)}{2} \right| > \lambda \right\}$.

§24. Проверка независимости

1. Критерий независимости в случае нормальной модели, основанный на эмпирическом коэффициенте корреляции.

В нашей модели выборка имеет вид $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, где x_i — независимые нормальные случайные величины $N(m_x, \sigma_x)$, y_i — независимые нормальные случайные величины $N(m_y, \sigma_y)$, мы проверяем гипотезу о независимости x -ов и y -ов между собой, то есть о равенстве нулю коэффициента корреляции. Напомню формулу для вычисления эмпирического коэффициента корреляции между переменными x и y (после очевидного сокращения):

$$r = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \sqrt{\sum_{i=1}^n y_i^2 - n\bar{y}^2}}.$$

Итак, мы ищем распределение случайной величины r в предположении, что векторы (x_i) и (y_i) независимы, и компоненты каждого из них одинаково распределены и независимы. Мы сначала сведем эту случайную величину к виду

$$r = \frac{\xi_2}{\sqrt{\sum_{i=2}^n \xi_i^2}}, \quad (2)$$

где ξ_i независимы и нормальны $N(0, 1)$. В свою очередь, легко проверяется, что случайная величина вида (2) сводится к случайной величине с распределением Стьюдента преобразованием $t_{n-2} = \sqrt{n-2} \frac{r}{\sqrt{1-r^2}}$. Итак, по таблицам распределения t_{n-2} мы строим критерий в виде $\{|t_{n-2}| > \lambda\}$ и разрешаем это неравенство относительно r , получаем критерий вида $\{|r| > \lambda_1\}$.

Теперь докажем что случайные величины в левой и правой частях (2) имеют одно и то же распределение.

i) Заменой $\xi_i = \frac{x_i - m_x}{\sigma_x}$ и $\eta_i = \frac{y_i - m_y}{\sigma_y}$ мы приходим к представлению

$$r = \frac{\sum_{i=1}^n \xi_i \eta_i - n\bar{\xi}\bar{\eta}}{\sqrt{\sum_{i=1}^n \xi_i^2 - n\bar{\xi}^2} \sqrt{\sum_{i=1}^n \eta_i^2 - n\bar{\eta}^2}}. \quad (3)$$

Из этого представления уже видно, что распределение r не зависит от неизвестных нам параметров $m_x, m_y, \sigma_x, \sigma_y$. Дальше мы хотим избавиться от $\bar{\xi}$ и $\bar{\eta}$ в (3). Это опять достигается вращением на ортогональную матрицу U , первый столбец которой имеет вид $1/\sqrt{n}, \dots, 1/\sqrt{n}$. Мы вводим случайные вектор-строки $\vec{\varphi} = (\varphi_1, \dots, \varphi_n)$ ($\vec{\varphi}'$ — вектор-столбец), $\vec{\xi} = (\xi_1, \dots, \xi_n)$, $\vec{\psi} = (\psi_1, \dots, \psi_n)$, $\vec{\eta} = (\eta_1, \dots, \eta_n)$, которые связаны соотношениями $\vec{\varphi} = \vec{\xi}U$, $\vec{\psi} = \vec{\eta}U$. Имеем: $\sqrt{n}\bar{\xi} = \varphi_1$, $\sqrt{n}\bar{\eta} = \psi_1$. По правилам умножения матриц имеем также

$$\sum_{i=1}^n \xi_i \eta_i = \vec{\xi} \vec{\eta}' = \vec{\xi} U(\vec{\eta}' U)' = \vec{\varphi} \vec{\psi}' = \sum_{i=1}^n \varphi_i \psi_i.$$

Поэтому

$$\sum_{i=1}^n \xi_i \eta_i - n\bar{\xi}\bar{\eta} = \sum_{i=2}^n \varphi_i \psi_i.$$

Аналогично (и как и в выводах для доверительных интервалов), имеем:

$$r = \frac{\sum_{i=2}^n \varphi_i \psi_i}{\sqrt{\sum_{i=2}^n \varphi_i^2} \sqrt{\sum_{i=2}^n \psi_i^2}}. \quad (4)$$

ii) Теперь мы приведем (4) к виду (2). В терминах плотностей нам надо доказать, что

$$p_r(u) = p \frac{\xi_2}{\sqrt{\sum_{i=2}^n \xi_i^2}} (u). \quad (5)$$

Мы используем формулу для условной плотности (аналог формулы полной вероятности для плотностей)

$$p_r(u) = \int_{\mathbf{R}^{n-1}} p_r(u | \varphi_2 = v_2, \dots, \varphi_n = v_n) p_{\varphi_2, \dots, \varphi_n}(v_2, \dots, v_n) dv_2 \dots dv_n. \quad (6)$$

Дело в том, что условную плотность $p_r(u | \varphi_2 = v_2, \dots, \varphi_n = v_n)$ найти очень просто. Мы представляем вместо случайных величин $\varphi_2, \dots, \varphi_n$ их значения v_2, \dots, v_n и получаем случайную величину

$$r(v_2, \dots, v_n) = \frac{\sum_{i=2}^n v_i \psi_i}{\sqrt{\sum_{i=2}^n v_i^2} \sqrt{\sum_{i=2}^n \psi_i^2}} = \frac{\sum_{i=2}^n \lambda_i \psi_i}{\sqrt{\sum_{i=2}^n \psi_i^2}},$$

где

$$\lambda_i = \frac{v_i}{\sqrt{\sum_{i=2}^n v_i^2}}, \sum_{i=2}^n \lambda_i^2 = 1.$$

Итак,

$$p_r(u | \varphi_2 = v_2, \dots, \varphi_n = v_n) = p_{r(v_2, \dots, v_n)}(u).$$

Чтобы найти функцию плотности $p_{r(v_2, \dots, v_n)}(u)$, мы проведем вращение случайного вектора $U(\psi_i) = (\xi_2)$ так, что первая строка (точнее, строка с номером 2) ортогональной матрицы u совпадает с вектором (λ_i) . После вращения случайная величина $r(v_2, \dots, v_n)$ приводится к виду

$$r(v_2, \dots, v_n) = \frac{\xi_2}{\sqrt{\sum_{i=2}^n \xi_i^2}}.$$

Заметим, что плотность этой случайной величины не зависит от выбора (v_2, \dots, v_n) . Поэтому в интеграле (6) первый множитель постоянен и равен $p \frac{\xi_2}{\sqrt{\sum_{i=2}^n \xi_i^2}} (u)$, а интеграл по второму множителю (плотность!) равен 1. Мы доказали (5).

iii)

$$\sqrt{n-2} \frac{r}{\sqrt{1-r^2}} = \frac{\xi_2}{\sqrt{\frac{1}{n-2} \sum_{i=3}^n \xi_i^2}} = t_{n-2}. \quad (7)$$

Как все это используется. По таблицам распределения Стьюдента находим такое число ε , что

$$\mathbb{P}\{|t_{n-2}| > \varepsilon\} = \alpha,$$

где α — запланированный нами уровень значимости. Далее решаем уравнение (7) относительно r , то есть находим такое δ , что

$$|t_{n-2}| = \varepsilon \Leftrightarrow |r| = \delta.$$

После этого вычисляем эмпирический коэффициент корреляции r и проверяем неравенство $|r| > \delta$.

Замечание. Парадокс теории проверки гипотез. Допустим, что мы имеем не одну, а много одинаковых выборок, например, 500 (мы разбили одну большую выборку на много средних). Сделано это для того, чтобы быть более уверенными в справедливости гипотезы. Итак, мы проверяем 500 раз гипотезу с уровнем значимости 0.05, и все 500 раз подтверждается гипотеза. О чём это говорит? О том, что гипотеза неверна. Если бы она была верна, то примерно 25 раз гипотеза должна была бы не подтвердиться.

Когда я рассказал этот парадокс одному очень умному практику и его сотрудникам, он сказал: 'Не слушайте его, это означает, что подтверждается лучшая гипотеза'. И это разумно, но, как мы сейчас увидим, не всегда. Контрпример – проверка на независимость в рамках нормальной модели. Если гипотеза независимости все время подтверждается, то это указывает на ошибочность исходной нормальной модели. Но если совместное распределение не является нормальным, то даже равенство коэффициента корреляции нулю не влечет независимость.

2. Критерий независимости, основанный на методе χ^2 .

Если информации о нормальности совместного распределения у нас нет, то критерий независимости наблюдаемых случайных величин x и y может быть основан на методе χ^2 . Для этого мы должны разделить область значений x на s частей, $\Omega = A_1 + \dots + A_s$, а область значений y на t частей, $\Omega = B_1 + \dots + B_t$. Мы также обозначим:

n_{jk} – число попадания вектора (x, y) в $A_j \cap B_k$,

$$\sum_{j,k} n_{jk} = n, \sum_k n_{jk} = n_{\cdot j}, \sum_j n_{jk} = n_{\cdot k}.$$

Параметрами модели будут числа

$$p_{\cdot j} = \mathbb{P}(A_j), p_{\cdot k} = \mathbb{P}(B_k).$$

Так как суммы вероятностей по j и по k равны 1, общее число параметров равно $t + s - 2$. Проверяется гипотеза $\mathbb{P}(A_j B_k) = p_{\cdot j} p_{\cdot k}$. Оценки для параметров находятся с помощью упрощенной процедуры минимума χ^2 и имеют естественный вид: $p_{\cdot j} = n_{\cdot j}/n$, $p_{\cdot k} = n_{\cdot k}/n$. Итак, для проверки независимости мы имеем статистику

$$T = \sum_{j,k} \frac{(n_{jk} - n_{\cdot j} n_{\cdot k} / n)^2}{n_{\cdot j} n_{\cdot k} / n}$$

с распределением $\chi^2_{st-s-t+1}$.

3. Ранговые критерии применяются в ситуации, когда у нас нет разумной модели (часто это бывает в задачах психологии и педагогики), а сами величины наблюденных нами характеристик не имеют существенного значения, важны лишь неравенства между этими числами. Например, мы знаем, что оценка 5 лучше 4, а 4 лучше 3, но были бы очень странными утверждения '5 лучше 4 в 5/4 раза', '4 лучше 3 в 4/3'. (Впрочем, если бы стипендия была прямо пропорциональна сумме баллов, оценки превратились бы в рубли, и эти соотношения стали бы очень даже разумными). Итак, в такого рода ситуации целесообразно заменить наблюденные значения x_1, \dots, x_n на числа $\tilde{x}_1, \dots, \tilde{x}_n$ из множества $\{1, 2, \dots, n\}$ по правилу: $x_i < x_j$ влечет $\tilde{x}_i < \tilde{x}_j$ (для простоты мы будем рассматривать лишь случай, когда все значения чисел x_i различны). Аналогично мы заменяем числа y_1, \dots, y_n на числа $\tilde{y}_1, \dots, \tilde{y}_n$ из множества $\{1, 2, \dots, n\}$. Например, число 1 мы ставим на место самого маленького числа из всех x_i и самого маленького числа из всех y_j , а число n мы ставим на место самого большого числа из всех x_i и самого большого числа из всех y_j . Итак, мы имеем набор векторов $(\tilde{x}_1, \tilde{y}_1), \dots, (\tilde{x}_n, \tilde{y}_n)$. Теперь мы можем проверить независимость x -ов и y -ов. Для этого мы считаем эмпирический коэффициент корреляции \tilde{r} для двумерной выборки $(\tilde{x}_1, \tilde{y}_1), \dots, (\tilde{x}_n, \tilde{y}_n)$. Если число $|\tilde{r}|$ окажется велико, то мы можем сказать, что 'гипотеза независимости не подтвердилась'. Но мы должны уметь определять, что значит 'велико'. Гипотеза независимости означает, что все варианты наборов $(\tilde{x}_1, \tilde{y}_1), \dots, (\tilde{x}_n, \tilde{y}_n)$ равновероятны, то есть вероятность каждого набора в нашей модели равна

квадрату числа перестановок множества $\{1, 2, \dots, n\}$, то есть $(\frac{1}{n!})^2$. Для каждого такого набора легко считается свое $|\tilde{r}[(\tilde{x}_1, \tilde{y}_1), \dots, (\tilde{x}_n, \tilde{y}_n)]|$. Но как вычислить $\mathbb{P}\{\tilde{r} > \tilde{r}_0\}$? Аналитическое (даже приближенное) построение функции распределения \tilde{r} для конкретного n представляется очень сложной задачей. Для этого нужно упорядочить $n!$ чисел. Но вам эта функция распределения не нужна. Для вашего конкретного случая вы вычислили эмпирический коэффициент для рангов (обозначим результат \tilde{r}_0). Далее перебрали на компьютере все возможные элементарные исходы и для каждого из них подсчитали \tilde{r} , подсчитали долю исходов, для которых $\tilde{r} > \tilde{r}_0$ — это и будет нужная вам вероятность. Заметим, что общее число элементарных исходов можно считать равным не $n!$, а лишь $n!$. Это следует из соображений симметрии. Мы можем рассматривать лишь исходы вида $[(1, \tilde{y}_1), (2, \tilde{y}_2), \dots, (n, \tilde{y}_n)]$ и лишь для них подсчитать эмпирический коэффициент корреляции. Разумеется, нужна программа, перебирающая все перестановки множества $\{1, 2, \dots, n\}$.

Замечание. В русском переводе книги 'M. Kendall. Rank Correlation Methods' (М. Кендалл. Ранговые корреляции. М. Статистика. 1975) имеется таблица для $n = 10$, но разбираться в книге не очень удобно — слишком много текста, книга написана для неспециалистов. Как мне кажется,

$$\begin{aligned}\mathbb{P}\{|\tilde{r}| \geq \frac{91}{165}\} &\approx 0.104, \quad \mathbb{P}\{|\tilde{r}| \geq \frac{93}{165}\} \approx 0.096, \\ \mathbb{P}\{|\tilde{r}| \geq \frac{105}{165}\} &\approx 0.054, \quad \mathbb{P}\{|\tilde{r}| \geq \frac{107}{165}\} \approx 0.048,\end{aligned}$$

но неплохо было бы проверить с помощью компьютерных вычислений — правильно ли я разобрался в таблицах, или самим разобраться в этой книге или в других книгах. Определение используемого в таблице термина 'функция вероятности' я не нашел. Впрочем, компьютерные вычисления лучше со всех точек зрения, так как они дают точное значение уровня значимости для данной выборки (не обязательно, рядом с 0.1 или 0.05).

4. Коэффициент конкордации.

Для измерений степени тесноты статистической связи у более чем двух порядковых переменных $r > 2$ используется *коэффициент конкордации* Кендалла

$$\frac{12}{r^2(n^3 - n)} \sum_{i=1}^n \left(\sum_{j=1}^r R_i(j) - \frac{r(n+1)}{2} \right)^2,$$

где $R_i(j)$ — ранг i -го наблюдения j -й случайной величины, r — число переменных, n — число наблюдений.

5. Проверка нормальности.

Проверить гипотезу нормальности генеральной совокупности, из которой извлечена данная одномерная выборка (x_1, \dots, x_n) , можно, например, с помощью критерия χ^2 . Для этого надо разделить числовую прямую на r частей, оценивание параметров t и σ^2 проводить методом минимума χ^2 . Другие критерии используют свойства моментов стандартного нормального распределения. А именно, если ξ нормальна $N(0, 1)$, то $\mathbb{E}\xi^3 = 0$. Третий центральный нормированный момент случайной величины называется *асимметрией*. Эмпирическая асимметрия имеет вид

$$A = \frac{\frac{1}{n} \sum_{i=1}^n [x_i - \bar{x}]^3}{\left(\frac{1}{n} \sum_{i=1}^n [x_i - \bar{x}]^2 \right)^{3/2}}.$$

Легко показывается, как мы уже делали неоднократно, что в случае нормальной модели распределение эмпирической асимметрии не зависит от параметров t и σ . Соответствующие таблицы для распределения A имеются, используя их, вычисляется вероятность критического

множества $\{|A| > \varepsilon\}$. Известно также, что четвертый центральный нормированный момент нормального распределения равен 3. Характеристика $\frac{\mathbb{E}(\xi - \mathbb{E}\xi)^4}{[\mathbb{E}(\xi - \mathbb{E}\xi)^2]^2} - 3$ называется *экссесом* распределения. Эмпирический эксцесс также используется для построения критерия нормальности.

6. Вывод распределения Фишера. Критерий Фишера.

В классической ситуации сравнения двух нормальных выборок $(x_i)_{i=1}^n$ и $(y_j)_{j=1}^k$ целесообразно использовать с маленьким уровнем значимости α кроме самого критерия Стьюдента еще проверку на нормальность и *критерий Фишера*, позволяющий проверить гипотезу о равенстве дисперсий обеих выборок. Этот критерий основан на статистике Фишера

$$\varphi = \frac{\frac{1}{n}\chi_{n,x}^2}{\frac{1}{k}\chi_{k,y}^2},$$

где

$$\chi_{n,x}^2 = \sum_i (x_i - \bar{x})^2, \chi_{k,y}^2 = \sum_j (y_j - \bar{y})^2,$$

Очевидно, что φ сходится по вероятности к 1 (когда $n, k \rightarrow \infty$), поэтому критическое множество Фишера состоит из значений φ , существенно отклоняющихся от 1, где понятие 'существенно' определяется с помощью распределения Фишера статистики φ . Нам будет удобнее найти плотность распределения случайной величины $\psi = \frac{n}{k}\varphi$, выразить через нее плотность φ читатель должен самостоятельно.

Функцию распределения ψ обозначим через $F_{n,k}$. Мы используем стандартную для отношения двух независимых случайных величин замену: $y = u/v, z = v$. Имеем:

$$\begin{aligned} F_{n,k}(x) &= \int_{u/v \leq x, u \geq 0, v \geq 0} \int p_{\chi_n^2}(u)p_{\chi_k^2}(v)dudv = \\ &= \int_{u/v \leq x, u \geq 0, v \geq 0} \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} e^{-\frac{u}{2}} u^{\frac{n}{2}-1} I_{(0,\infty)}(u) e^{-\frac{v}{2}} v^{\frac{k}{2}-1} I_{(0,\infty)}(v)dudv = \\ &= \int_{0 \leq y \leq x, z \geq 0} \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} e^{-\frac{yz}{2}} (yz)^{\frac{n}{2}-1} \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} e^{-\frac{z}{2}} z^{\frac{k}{2}-1} dydz = \\ &= \int_0^x \frac{1}{2^{\frac{n+k}{2}} \Gamma(\frac{n}{2}) \Gamma(\frac{k}{2})} y^{\frac{n}{2}-1} dy \int_0^\infty e^{-(\frac{y+1}{2})z} z^{\frac{n+k}{2}-1} dz = \\ &= \int_0^x \frac{\Gamma(\frac{n+k}{2})}{\Gamma(\frac{n}{2}) \Gamma(\frac{k}{2})} \frac{y^{\frac{n}{2}-1}}{(y+1)^{\frac{n+k}{2}}} dy. \end{aligned}$$

Отсюда немедленно получается функция плотности.

§25. Различение двух гипотез методом последовательного анализа Вальда

В отличие от всего предыдущего, в этом разделе число наблюдений не является постоянным n , а является случайной величиной. Рассматривается задача различения двух гипотез: гипотезы H_0 , состоящей в том, что случайная величина имеет распределение \mathbb{P}_0 с функцией плотности $p_0(x)$, и гипотезы H_1 , состоящей в том, что случайная величина имеет распределение \mathbb{P}_1 с функцией плотности $p_1(x)$. На каждом шаге n для полученной выборки (x_1, \dots, x_n) вычисляется отношение $p_0^{(n)}/p_1^{(n)} = p_0(x_1) \cdots p_0(x_n)/p_1(x_1) \cdots p_1(x_n)$ двух совместных плотностей. Если $p_0^{(n)}/p_1^{(n)} > C_0$, то выбирается гипотеза H_0 , если $p_0^{(n)}/p_1^{(n)} < C_1$, то выбирается гипотеза

H_1 , если $C_1 \leq p_0^{(n)}/p_1^{(n)} \leq C_0$, то наблюдения продолжаются. Задача состоит в подборе чисел α и β при фиксированных вероятностях ошибок первого и второго рода α и β . Оказывается, числа α и β вычисляются только по α и β и не зависят от распределений \mathbb{P}_0 и \mathbb{P}_1 (если пренебречь ошибками при подсчете вероятностей, которые получаются из за перескока за указанные выше граничные).

В этом подходе области Δ_0 и Δ_1 являются суммами соответствующих областей для различных n :

$$\Delta_0 = \sum_n \Delta_0^{(n)}, \quad \Delta_1 = \sum_n \Delta_1^{(n)},$$

где $\Delta_0^{(n)}$ — событие, состоящее в принятии гипотезы H_0 на n -ом шаге, $\Delta_1^{(n)}$ — событие, состоящее в принятии гипотезы H_1 на n -ом шаге. Тогда

$$\alpha = \sum_n \mathbb{P}_0(\Delta_1^{(n)}), \quad \beta = \sum_n \mathbb{P}_1(\Delta_0^{(n)}).$$

По построению имеем для каждого n :

$$\mathbb{P}_0(\Delta_0^{(n)}) = \int_{\Delta_0^{(n)}} p_0^{(n)} \geq \int_{\Delta_0^{(n)}} C_0 p_1^{(n)} = C_0 \mathbb{P}_1(\Delta_0^{(n)})$$

$$\mathbb{P}_1(\Delta_1^{(n)}) = \int_{\Delta_1^{(n)}} p_1^{(n)} \geq \int_{\Delta_1^{(n)}} \frac{1}{C_1} p_0^{(n)} = \frac{1}{C_1} \mathbb{P}_0(\Delta_1^{(n)}).$$

Суммируя эти неравенства, получаем

$$1 - \alpha = \mathbb{P}_0(\Delta_0) \geq C_0 \mathbb{P}_1(\Delta_0) = C_0 \beta, \quad 1 - \beta = \mathbb{P}_1(\Delta_1) \geq \frac{1}{C_1} \mathbb{P}_0(\Delta_1) = \frac{1}{C_1} \alpha.$$

Итак,

$$C_0 \leq \frac{1 - \alpha}{\beta}, \quad C_1 \geq \frac{\alpha}{1 - \beta}.$$

§26. Равномерно наиболее мощные критерии

1. Определение. Пусть проверяется гипотеза H_0 (отвечающая распределению P_0) со сложной альтернативой $\mathcal{P}_1 = \{P_{1,\theta}\}_{\theta \in \Theta}$, критерий Δ таков, что

- i) уровень значимости критерия равен α ,
- ii) для любого другого критерия Δ' с тем же уровнем значимости

$$\alpha = P_0 \Delta = P_0 \Delta',$$

справедливо неравенство

$$\beta(\theta) = P_{1,\theta} \Delta \leq P_{1,\theta} \Delta' \text{ для всех } \theta.$$

Тогда критерий Δ называется *равномерно наиболее мощным*.

Мы приведем простой пример, когда действительно существует равномерно наиболее мощный критерий, лучший других при всех альтернативах $P_{1,\theta}$ и много примеров, когда такого критерия не существует. Заметим, что лемма Неймана — Пирсона однозначно определяет такой критерий (если он существует).

Пример 1. H_0 — выборка принадлежит нормальной генеральной совокупности $N(m_0, \sigma_0)$. Альтернативой является набор распределений $N(m, \sigma_0)$, где $m > m_0$. В этом случае критерий вида $\Delta = \{\bar{x} > C\}$ является равномерно наиболее общим. Это утверждение следует из леммы

Неймана—Пирсона. Действительно, для любого $m > m_0$ критерий, минимизирующий $\beta(m)$ при заданном α имеет вид

$$\begin{aligned} \Delta &= \left\{ \frac{p_{m_0, \sigma_0}^{(n)}(x_1, \dots, x_n)}{p_{m, \sigma_0}^{(n)}(x_1, \dots, x_n)} < C(m) \right\} = \\ &= \left\{ \frac{\left(\frac{1}{\sigma_0 \sqrt{2\pi}} \right)^n e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - m_0}{\sigma_0} \right)^2}}{\left(\frac{1}{\sigma_0 \sqrt{2\pi}} \right)^n e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - m}{\sigma_0} \right)^2}} < C(m) \right\} = \\ &= \left\{ -\sum_{i=1}^n \left(\frac{x_i - m_0}{\sigma_0} \right)^2 + \sum_{i=1}^n \left(\frac{x_i - m}{\sigma_0} \right)^2 < 2 \ln C(m) \right\} = \\ &= \left\{ 2 \sum_{i=1}^n x_i m_0 - 2 \sum_{i=1}^n x_i m < 2\sigma_0^2 \ln C(m) + nm_0^2 - nm^2 \right\}. \end{aligned}$$

Разделив это неравенство на $2(m_0 - m)n$ и обозначив правую часть через C , мы получим множество требуемого вида. Заметим, что C зависит только от α (и от m_0 и σ_0), но не зависит от m . Действительно, находится из равенства

$$\mathbb{P}_{m_0, \sigma_0}^{(n)}\{\bar{x} > C\} = \alpha.$$

Пример 2. H_0 — выборка принадлежит нормальной генеральной совокупности $N(m_0, \sigma_0)$. Альтернативой является набор распределений $N(m, \sigma_0)$, где $m \neq m_0$. Действуя так же, мы получим множество вида $\Delta = \{\bar{x} > C\}$ для альтернатив $m > m_0$, но множество вида $\Delta = \{\bar{x} < C\}$ для альтернатив $m < m_0$. К сожалению, одно множество, которое минимизировало бы $\beta(m)$ и для $m > m_0$ и для $m < m_0$, мы получить не можем.

Замечание. В обеих задачах ничего хорошего не получается при неизвестной дисперсии, так как константа C , определяющая множество Δ по уровню значимости α зависит от параметра σ .

2. Дополнение. Слова.

Если мы попали в критическое множество, то некорректно говорить, что наша гипотеза *неверна*. Может так оказаться, что на самом деле гипотеза верна, но нам не повезло. Поэтому правильнее сказать, что гипотеза *не подтвердилась*, а если мы не попали в критическое множество, то гипотеза *подтвердилаась*. В ситуации, когда одновременно с основной гипотезой мы рассматриваем альтернативную гипотезу, не совсем корректно утверждать при попадании в критическое множество, что *подтвердилаась альтернативная гипотеза*. Дело в том, что (малую!) вероятность критического множества мы вычислили лишь в предположении справедливости основной гипотезы. Такое высказывание оказывается еще менее разумным, если альтернативная гипотеза является сложной, тогда вероятность критического множества зависит от выбора параметра в рамках альтернативной гипотезы. В то же время в условиях леммы Неймана — Пирсона такого рода утверждение может оказаться разумным, но тогда вероятность β также должна быть мала, и критические множества для обеих гипотез, основной и альтернативной, должны являться дополнениями друг друга.

3. Дополнение.

Напомним определение условной плотности

$$p_\xi(x|\eta = y) = \frac{p_{\xi,\eta}(x, y)}{\int_{-\infty}^{\infty} p_{\xi,\eta}(x, y) dx}.$$

Понятие условной плотности само является условным, так как изменение значения совместной плотности на множестве нулевой меры (например, на прямой $\{x = x_0\}$) не меняет совместного распределения, но полностью меняет условную плотность. Однако рассмотрим случайную величину $f(\xi, \eta)$ и вычислим ее плотность следующим образом:

$$p_{f(\xi, \eta)}(x) = \int_{-\infty}^{\infty} p_{f(\xi, y)}(x)p_{\eta}(y)dy. \quad (3)$$

Нетрудно видеть, что определенная так плотность действительно дает нам распределение случайной величины $f(\xi, \eta)$. Действительно, по определению плотности мы должны иметь

$$\mathbb{P}\{f(\xi, \eta) \in B\} = \int_B p_{f(\xi, \eta)}(x)dx.$$

Подставим в это равенства вместо $p_{f(\xi, \eta)}$ выражение из (3) и получим тождество по теореме Фубини:

$$\begin{aligned} \int_B dx \int_{-\infty}^{\infty} p_{f(\xi, y)}(x)dy &= \int_{-\infty}^{\infty} p_{\eta}(y)dy \int_B p_{f(\xi, y)}(x)dx = \\ &= \int_{-\infty}^{\infty} p_{\eta}(y)dy \int_B p_{f(\xi, y)}(x)dx \end{aligned}$$

Критерий Колмогорова-Смирнова. Критерий Спирмена. Является ли оценка максимального правдоподобия для параметра t логнормального распределения эффективной? Область безразличия.

§27. Многомерный анализ (обзор)

Здесь мы не будем приводить каких-либо алгоритмы и формулы, а напишем лишь об известных в литературе постановках задач и подходах к их решениям. Более подробно с этим материалом можно познакомиться, например, в книгах [11], [32].

Наблюдается n раз r -мерный случайный вектор. Результаты наблюдения составляют выборку $(x_1^{(1)}, \dots, x_r^{(1)}), \dots, (x_1^{(n)}, \dots, x_r^{(n)})$. По этим данным мы хотим получить некую информацию о связи координат наблюдаемого вектора. Очевидно, что n должно быть не меньше r , иначе координаты x_i , которые являются векторами в n -мерном пространстве, связаны между собой линейными соотношениями бесконечным числом способов. Для разумного статистического вывода нужно, чтобы n было существенно больше r . Люди с гуманитарным образованием часто пренебрегают этим требованием.

1. Регрессия.

Представляет большой практический интерес задача о наилучшем приближении значения одной координаты, x_1 , функцией $f(x_2, \dots, x_r)$ других координат. Мы используем метод наименьших квадратов, а именно, минимизируем число $\sum(x_1 - f(x_2, \dots, x_r))^2$, где суммирование производится по всем полученным в результате эксперимента выборкам. Заметим, что возможна ситуация, когда ошибка эксперимента зависит от значения x_1 , тогда слагаемые в сумме надо умножать на некоторые веса (если ошибка больше, вес наблюдения должен быть меньше). В курсе теории вероятностей мы решали аналогичную задачу наилучшего приближения случайной величины η функцией $f(\xi_1, \dots, \xi_r)$ других случайных величин. Под словом 'наилучшее' имелся в виду выбор такой функции f , для которой минимально число $\mathbb{E}(\eta - f(\xi_1, \dots, \xi_r))^2$. Решением этой задачи является условное математическое ожидание $\mathbb{E}(\eta|\xi_1, \dots, \xi_r)$. Обычно такое решение неприменимо в задачах математической статистики. Дело в том, что если наблюдения проводить достаточно точно, то набор (x_2, \dots, x_r) встречается лишь в одной выборке (x_1, x_2, \dots, x_r) и наилучшей функцией оказывается $f(x_2, \dots, x_r) = x_1$. Разумеется, такая 'функция' задана лишь на множестве результатов проведенных наблюдений и ничего не говорит о наблюдениях, которые еще не проведены. Поэтому в математической статистике нужно искусственно ограничивать класс функций f и решать задачу регрессии лишь в этом классе.

Возникает произвол в выборе класса, функции оказываются зависимыми от некоторого конечного числа параметров, и наша задача сводится к задаче нахождения минимума функции нескольких переменных. Например, если мы ищем функцию f в классе всех линейных функций

$$l(x_2, \dots, x_r) = a_2x_2 + \dots + a_rx_r + b,$$

мы должны решить систему уравнений

$$\begin{aligned} \frac{\partial}{\partial a_2} \sum_{i=1}^n (x_1^{(i)} - [a_2x_2^{(i)} + \dots + a_rx_r^{(i)} + b])^2 &= 0, \\ \dots \\ \frac{\partial}{\partial a_r} \sum_{i=1}^n (x_1^{(i)} - [a_2x_2^{(i)} + \dots + a_rx_r^{(i)} + b])^2 &= 0, \\ \frac{\partial}{\partial b} \sum_{i=1}^n (x_1^{(i)} - [a_2x_2^{(i)} + \dots + a_rx_r^{(i)} + b])^2 &= 0. \end{aligned}$$

Не всегда разумно искать f в классе линейных функций. Например, в случае, когда наблюдения $y^{(i)}$ мы стремимся приблизить функцией f наблюдений $x^{(i)}$, может оказаться правильным выбрать функцию $f(x) = e^{ax+b}$ и минимизировать $\sum_{i=1}^n (y^{(i)} - e^{ax^{(i)}+b})^2$. Однако удобнее использовать линейную зависимость для логарифма y , но тогда мы придем к другому решению. Наконец, если возможна периодическая зависимость, целесообразно искать f в виде $a \sin(x + b) + c$.

Замечание для практического применения. Гаусс, который предложил метод наименьших квадратов, по-видимому, использовал его для уточнений траекторий (значительную часть жизни он был директором астрономической обсерватории). Наблюдения одним человеком проводились в разные моменты времени, по-существу, наблюдался случайный процесс — реальная кривая, зависящая от нескольких параметров, портилась случайным процессом помех. Для оценки параметров использовался изложенный выше метод, но только выше $x(t_i)$, $y(t_i)$, $z(t_i)$ приближаются результатами наблюдений x_i , y_i , z_i и нужно найти параметры, для которых сумма квадратов ошибок минимальна. По-видимому, Гаусс заменил координаты в пространстве на расстояния на небесной сфере, и через эти числа выражал сумму квадратов. Метод наименьших квадратов можно применять для анализа случайных процессов, например, при анализе изменения курса валют, но очень осторожно ввиду зависимости этого процесса от правительственные решений.

2. Корреляция.

Кроме выборочного коэффициента корреляции (для которого в модели двух независимых нормальных случайных величин нам удалось получить распределение) для векторных выборок размерности больше 2 рассматриваются две другие выборочные характеристики, которые могут дать дополнительное представление об изучаемом объекте. Здесь мы дадим лишь геометрическое объяснение для этих характеристик, вывод формул требует выкладок с матрицами. Итак, мы имеем выборку $(x_1^{(1)}, \dots, x_r^{(1)}), \dots, (x_1^{(n)}, \dots, x_r^{(n)})$ длины n (наблюдений) из r -мерных векторов (параметров) и матрицу $[r_{ij}]_{i,j \leq r}$ выборочных коэффициентов корреляции (диагональ которой состоит из единиц). Таким образом, в пространстве параметров \mathbf{R}^r определено скалярное произведение

$$A([x_i]_{i \leq r}, [y_i]_{i \leq r}) = \sum_{i,j=1}^n r_{ij}x_iy_j.$$

Сопоставляя i -му параметру вектор $e_i = [0, \dots, 0, 1, 0, \dots]$ (1 на i -м месте, остальные координаты равны 0), мы имеем: $r_{ij} = A(e_i, e_j)$. Напомним, что ортогональное проектирование P на линейное подпространство L переводит элемент x линейного пространства в такой элемент $Px \in L$, что $A(x - Px, y) = 0$ для всех $y \in L$. Разумеется, для проверки этого свойства достаточно проверить его для всех y из некоторого базиса в L . Теперь введем *частный* (выборочный)

коэффициент корреляции элементов e_1 и e_2 (для любых других e_i, e_j он определяется аналогично). Рассмотрим проекции \tilde{e}_1 и \tilde{e}_2 элементов e_1 и e_2 на линейное подпространство, порожденное всеми векторами e_3, \dots, e_r . Тогда частный коэффициент корреляции 1 и 2 координат обозначается $r_{12.34\dots r}$. Имеем:

$$r_{12.34\dots r} := \frac{A(e_1 - \tilde{e}_1, e_2 - \tilde{e}_2)}{A(e_1 - \tilde{e}_1, e_1 - \tilde{e}_1)^{1/2} A(e_2 - \tilde{e}_2, e_2 - \tilde{e}_2)^{1/2}}.$$

Частный коэффициент корреляции 1 и 2 координат можно интерпретировать как характеристику зависимости этих координат после того, как убрали из них зависимости от других координат, которые также связывают эти две координаты. Действительно, если x и y зависят от одного и того же z , то очевидно они зависят друг от друга. При вычислении частного коэффициента корреляции мы эти зависимости убираем.

Замечание. Из приведенной выше интерпретации не следует, что $r_{12.34\dots r}$ меньше по абсолютной величине чем r_{12} . Пример для случайных величин, тогда коэффициент корреляции обозначается буквой ρ . Пусть ξ, η независимы и нормальны с параметрами $N(0, 1)$. Рассмотрим $x_1 = \xi + \eta$, $x_2 = \xi - \eta$, $x_3 = \xi$. Тогда

$$\rho_{12} = 0, \quad \rho_{12.3} = -1.$$

Сводный коэффициент корреляции. Рассмотрим проекцию Pe_1 элемента e_1 на линейное пространство L , порожденное элементами e_2, \dots, e_r . Сводный коэффициент корреляции элемента e_1 и всех элементов e_2, \dots, e_r определяется равенством

$$r_{1(2\dots r)} := \frac{A(e_1, Pe_1)}{A(e_1, e_1)^{1/2} A(Pe_1, Pe_1)^{1/2}}.$$

Замечание. Как обычно, будем считать координаты $x_k^{(\cdot)}$ ($i \leq n$) векторов выборки случайными величинами, заданными на вероятностном пространстве из n элементарных исходов с вероятностью $1/n$, каждый элементарный исход — элемент выборки. В этой ситуации будем решать задачу регрессии, аппроксимируя случайную величину $x_1^{(\cdot)}$ линейной комбинацией $l(x_2^{(\cdot)}, \dots, x_r^{(\cdot)})$ случайных величин $x_2^{(\cdot)}, \dots, x_r^{(\cdot)}$ или аппроксимируя случайные величины $x_1^{(\cdot)}$ и $x_2^{(\cdot)}$ линейными комбинациями $l_1(x_3^{(\cdot)}, \dots, x_r^{(\cdot)})$ и $l_2(x_3^{(\cdot)}, \dots, x_r^{(\cdot)})$ случайных величин $x_3^{(\cdot)}, \dots, x_r^{(\cdot)}$. Тогда частный коэффициент корреляции $r_{12.34\dots r}$ равен коэффициенту корреляции случайных величин $x_1^{(\cdot)} - l_1(x_3^{(\cdot)}, \dots, x_r^{(\cdot)})$ и $x_2^{(\cdot)} - l_2(x_3^{(\cdot)}, \dots, x_r^{(\cdot)})$, а сводный коэффициент корреляции $r_{1(2\dots r)}$ равен коэффициенту корреляции случайных величин $x_1^{(\cdot)}$ и $l(x_2^{(\cdot)}, \dots, x_r^{(\cdot)})$.

3. Факторный анализ

С помощью метода факторного анализа пытаются уменьшить число координат наблюдаемого вектора. А именно, делается попытка выявить скрытые факторы, влияющие не результаты экспериментов. В используемой вероятностной модели предполагается, что случайный вектор $(x_i)_{i \leq r}$ наблюдаемых представим в следующем виде:

$$x_i = \sum_{j=1}^d l_{ij} z_j + e_i,$$

где $d < r$ (d не должно быть большим), случайные величины z_j независимы, нормальны, центрированы и имеют дисперсию 1, величины e_i представляют собой шумы при наблюдениях, поэтому они независимы между собой и с z_j , нормальны. Задача состоит в наилучшей оценке коэффициентов l_{ij} . Зная эти коэффициенты, мы можем выразить z_j через x_i , эти новые величины называются факторами. Как видите, в этом подходе число факторов задается заранее, от выбора этого числа зависит результат.

Решение осуществляется методом максимального правдоподобия, то есть записывается совместная плотность n наблюдений вектора (x_i) , которая зависит от параметров l_{ij} и дисперсий

и средних величин e_i , и вычисляется максимум (по неизвестным параметрам) этой плотности для данных результатов наблюдений. В результате мы находим l_{ij} и дисперсии e_i . Легко видеть, что факторы восстанавливаются по матрице $[l_{ij}]$ неоднозначно, с точностью до вращения. Практики (психологи и социологи) стараются с помощью такого вращения придать факторам разумный смысл, например, линейная комбинация координат, отвечающих за трудолюбие, математические способности, гуманитарные способности, память, выносливость, финансирование и т. д. (см. [32]). Далее результатам придается гуманитарное звучание.

Для корректности нужно провести проверку самой модели с помощью какого-нибудь критерия. Если в процессе проверки мы попали в критическое множество, то нужно увеличить число факторов в модели.

4. Метод главных компонент.

Метод главных компонент также служит снижению размерности и выявлению факторов. Использование этого метода имеет особенность в случае, когда все координаты можно сравнить друг с другом (например, время, затраченное на одно, другое и т.д.). Тогда мы рассматриваем матрицу вторых центральных моментов и приводим ее к главным осям. Векторы, соответствующие самым большим собственным векторам, называются главными компонентами, остальные считаются помехами. Чтобы вспомнить процедуру вычисления собственных чисел, надо посмотреть курс линейной алгебры.

Немного отличная ситуация возникает тогда, когда координаты несравнимы между собой (например, килограммы и метры). Тогда обычно наблюдаемые величины нормируются и матрица смешанных центральных моментов заменяется на матрицу коэффициентов корреляции.

Заметим, что и в этой ситуации следует проверить незначимость проигнорированных нами маленьких собственных значений. Для такой проверки разрабатываются критерии.

Часть 5. СЛУЧАЙНЫЕ ПРОЦЕССЫ

Введение

Под *случайным процессом* понимается коллекция случайных величин, заданных на одном вероятностном пространстве $(\Omega, \mathcal{A}, \mathbb{P})$ — (ξ_t) , где либо $t \geq 0$, либо $t \in \mathbf{R}$, либо $t \in T \subset \mathbf{R}$ (*процесс с непрерывным временем*), (ξ_n) , $n \in \{0\} \cup \mathbf{N}$ или $n \in \mathbf{Z}$ (*процесс с дискретным временем*). Случайные процессы бывают одномерными или многомерными (например, реальный процесс броуновского движения). Более общее понятие *случайной функции* ξ , индексированной точками произвольной природы, включает в себя также важнейшие для анализа изображений случайные поля.

Изучение случайного процесса начинается с рассмотрения трех определяемых в терминах процесса объектов:

- i) *среднее процесса* $m_\xi(t) = \mathbb{E}\xi_t$, $t \in T$;
- ii) *ковариация процесса* $r_\xi(t, s) = \mathbb{E}(\xi_t - \mathbb{E}\xi_t)(\xi_s - \mathbb{E}\xi_s)$, $t, s \in T$.
- iii) *совместные распределения процесса* $\mathbb{P}_{\xi_{t_1} \dots \xi_{t_n}}$, где $\{t_1, \dots, t_n\}$ — произвольное конечное подмножество области определения процесса T ;

Разумеется, среднее и ковариация процесса существуют не всегда, примером является класс *процессов Леви*, которые в последнее время все чаще используются при моделировании процессов в экономике. Впрочем, наболее часто используются *гауссовые процессы*, у которых все совместные распределения нормальны. Как мы уже видели, совместные распределения такого процесса однозначно определяются средним и ковариацией процесса.

Во многих ситуациях, и мы с этим неоднократно столкнемся, процесс задается не как зависящая от времени функция на вероятностном пространстве, а набором своих совместных распределений. Знаменитая теорема Колмогорова доказывает, что в этом случае можно придумать вероятностное пространство и процесс на нем, который будет иметь заданный набор совместных распределений.

Теория случайных процессов находит многочисленные применения: задачи прогнозирования, задача о разладке (определение момента внезапной смены одного наблюдаемого случайного процесса другим), процессы массового обслуживания, финансовая математика. Теория случайных процессов предоставляет модели для физики, механики, химии, биологии, метеорологии, практической статистики, гуманитарных наук.

Этот раздел теории вероятности использует многие глубокие факты теории меры, и, в свою очередь, стимулирует развитие этой теории. Он теснейшим образом связан с функциональным анализом. Важнейшие для приложений процессы второго порядка. в частности, стационарные случайные процессы, интерпретируются как кривые в гильбертовом пространстве. Другие важные процессы не имеют второго момента, развитие их теории связано с изучением других классов банаховых пространств. Случайные функции рассматриваются не только на евклидовых пространствах, но и на различных геометрических многообразиях. На основе теорий случайных процессов развивается важная для изучения процессов управления теория стохастических дифференциальных уравнений.

После этого введения придется ограничить задачи нашего курса лишь изучением начальных разделов этой теории и знакомством с рядом более продвинутых разделов.

Упражнение 1. Найти парные совместные распределения $\mathbb{P}_{\xi_t \xi_s}$, а также функции среднего и ковариации для процесса ξ_t , который задан соотношением

$$\xi_t(\omega) = I_{[0,t]}(\omega), t \in [0, 1], \omega \in [0, 1],$$

на вероятностном пространстве $[0, 1]$ с борелевской σ -алгеброй событий и вероятностной мерой \mathbb{P} , которая является обычной мерой Лебега, $T = [0, 1]$.

§28. Случайное блуждание.

Предварительные замечания

Один из способов вычисления вероятностей в классической модели интересующего нас события A — это представления его в виде суммы или разности событий, вероятности которых вычислить легче.

$$A = A_1 + A_2 + \dots + A_n$$

или

$$A = A_1 - A_2$$

так, что вероятности событий A_i легко вычисляются.

Иногда имеет смысл представить $\Omega = \Omega_1 \times \Omega_2$, каждый случайный исход при этом представляется $\omega = (\omega_1, \omega_2)$, где ω_1 — траектория ω до некоторого фиксированного момента k включительно, а ω_2 — траектория ω от момента $k+1$ до n . Иногда наше событие, которое нас интересует, можно представить в виде произведения $A = A_1 \times A_2$, где событие A_1 формулируется в терминах условий на траектории до момента k включительно, а A_2 в терминах условий на траектории от момента $k+1$ до n . Очевидно, что тогда $|A| = |A_1| \times |A_2|$.

1. Модель симметричного случайного блуждания.

Элементарный исход — траектория, состоящая из отрезков, направленных вверх \nearrow и направленных вниз \searrow . Эта модель по существу совпадает с моделью Бернулли, но некоторым событиям можно сопоставить новый геометрический смысл. Начинается траектория с момента времени 0 с точки 0 на числовой оси, то есть с точки $(0, 0)$ на координатной плоскости. Итак, если время движения равно n , то $|\Omega| = 2^n$. Ввиду симметричности движения вероятности всех элементов Ω равны 2^{-n} . Однако при изменении числа шагов случайного блуждания Ω меняется. Имеются задачи и для бесконечного числа шагов, тогда множество элементарных исходов оказывается бесконечным и даже несчетным, точнее, оно совпадает с множеством всех двоичных разложений чисел отрезка $[0, 1]$. Саму траекторию мы будем обозначать ω , а значение случайного блуждания в момент времени n через ξ_n .

Физики любят такие простые модели, на простых примерах они стараются уловить эффекты, которые могут произойти в более сложной ситуации.

Введем события $A_{n,k}$, состоящие из всех траекторий, кончающихся в точке (n, k) . $A_{n,k} = \{\xi_n = k\}$. Вычислим вероятность $\mathbb{P}(A_{n,k})$, которая равна $|A_{n,k}|/2^n$. Очевидно, что для попадания в точку (n, k) наша траектория должна содержать $\frac{n+k}{2}$ участков \nearrow и $\frac{n-k}{2}$ участков \searrow . Число таких траекторий мы можем отождествить с числом вариантов подбрасывания монеты, где $\frac{n+k}{2}$ гербов и $\frac{n-k}{2}$ решек. Оно равно $C_n^{\frac{n+k}{2}} = C_n^{\frac{n-k}{2}}$. Обратим внимание на одну особенность модели — в точке (n, k) числа n и k имеют одинаковую четность.

2. Задача о пьяном гуляке.

Случайное блуждание мы будем считать как движение одномерного пьяного гуляки. Его домом мы считаем точку 0, город кончается в точке r .

Задача. Найти вероятность того, что за n шагов пьяный гуляк хотя бы раз достигнет границы города. Другая, менее приятная версия этой задачи — такая. Пьяный гуляка лежит в канаве на расстоянии r шагов от оврага. Каждую минуту он делает шаг к оврагу или от него с вероятностью $\frac{1}{2}$. Если он достигнет оврага, то обязательно в него упадет. Через n минут за ним приедет машина. Какова вероятность, что пьяный гуляка упадет в овраг?

Замечу, что наше решение будет иметь вид не очень красивой формулы для подсчета вероятности.

Введем обозначения для событий:

$$A = \{\text{событие, состоящее в том, что траектория коснулась уровня } r\}$$

Тогда

$$A = \sum_{k \geq r} A_{n,k} + \sum_{k < r} AA_{n,k}.$$

Введем обозначение для $k < r$:

$$B_{n,k} = AB_{n,k} = \{\text{событие, состоящее в том, что траектория кончается в точке } (n, k), \text{ но по дороге касается уровня } r\}.$$

2. Лемма об отражении. $|B_{n,k}| = |A_{n,2r-k}|$.

Доказывается путем построения взаимно-однозначного соответствия между множествами $B_{n,k}$ и $A_{n,2r-k}$. Оно осуществляется следующим образом: до момента первого касания уровня r траекторию $\omega \in B_{n,k}$ мы оставляем неизменной, а после этого касания симметрично отображаем относительно прямой $y = r$.

Из этой леммы следует следующая формула для вычисления $\mathbb{P}(A)$:

$$\mathbb{P}(A) = \sum_{k \geq r} \mathbb{P}(A_{n,k}) + \sum_{k < r} \mathbb{P}(B_{n,k}) = \left\{ 2 \sum_{k > r} C_n^{\frac{n+k}{2}} + C_n^{\frac{n+r}{2}} \right\} 2^{-n}.$$

3. Независимость будущего от прошлого при фиксированном настоящем. Марковское свойство.

Марковское свойство процесса означает независимость будущего от прошлого при известном настоящем. Напомним, что события A и B называются *условно независимыми при условии C* , если

$$\mathbb{P}(AB|C) = \mathbb{P}(A|C)\mathbb{P}(B|C).$$

Для случайного процесса ξ_t это свойство должно выглядеть так: при $s < t < u$

$$\mathbb{P}(\{\xi_u = x, \xi_s = y\}|\xi_t = z) = \mathbb{P}(\xi_u = x|\xi_t = z)\mathbb{P}(\xi_s = y|\xi_t = z).$$

Однако если случайные величины ξ_u , ξ_t , ξ_s имеют непрерывный тип распределения, то все выписанные вероятности будут равны нулю, также вероятность условия будет равна нулю, а на нуль делить нельзя. Положение не спасает введение в рассмотрение событий вида $\{\xi_u \in A\}$, $\{\xi_s \in C\}$, $\{\xi_t \in B\}$. Действительно, если мы возьмем $B = \mathbf{R}$, то событие $\{\xi_t \in \mathbf{R}\}$ будет достоверным и из условной независимости будет вытекать обычная независимость.

Таким образом, введение понятия марковости требует использования тонких понятий условного среднего в самой общей ситуации. В нашем случае симметричного случайного блуждания проблем с корректным определением не возникает, так как

$$\mathbb{P}(\{\xi_n = k\}) = \frac{C_n^{\frac{n+k}{2}}}{2^n} \neq 0$$

при одинаковой четности n и k .

Упражнение. Независимость приращений случайного блуждания. Проверьте, что события $\{\xi_n - \xi_m = i\}$ и $\{\xi_k - \xi_l = j\}$ независимы при $l < k \leq m < n$. Для этого найдите число траекторий в каждом из этих событий и в их пересечении.

Из этого упражнения следует, что вероятность $\mathbb{P}(\xi_u = x)$ зависит от положения случайного блуждания в момент времени t , но не зависит от других свойств предшествующего пути.

Комментарий. Марковость случайного блуждания можно обосновать и по другому, более строго. Будем использовать сохранение независимости при операции сложения несовместных событий (см. пункт 4.7). Это полезно, так как удобно проверять независимость самых маленьких событий, являющихся элементарными исходами в соответствующих моделях.

Событие $C = \{ \text{набор шагов } \searrow \text{ и } \nearrow \text{ за время от } s \text{ до } t \}$.

Событие $D = \{ \text{набор шагов } \searrow \text{ и } \nearrow \text{ за время от } m \text{ до } n \}$.

$$\mathbb{P}(C) = 2^{-(t-s)}, \mathbb{P}(D) = 2^{-(n-m)}, \mathbb{P}(CD) = 2^{-[(t-s)+(n-m)]}.$$

В то же время случайные величины ξ_n и ξ_m при $m < n$, вообще говоря зависимы. Возьмем, например, события $\{\xi_{10} = 6\}$ и $\{\xi_{11} = 9\}$. Это события с ненулевой вероятностью, но они несовместны. Следовательно, они зависимы.

Докажем независимость будущего случайного блуждания от прошлого при фиксированном настоящем p (т. е. независимость его значений при $n \geq t > p$ от его значений при $s \leq p$ при условии $A_{p,r} = \{\xi_p = r\}$). Это свойство случайного блуждания называется *марковским*. Мы должны доказать равенство

$$\mathbb{P}(CD|A_{p,r}) = \mathbb{P}(C|A_{p,r})\mathbb{P}(D|A_{p,r}) \quad (1)$$

для любого события C , означающего некое условие на значения процесса в моменты времени $s \leq p$ и для любого события D , означающего некое условие на значения процесса в моменты времени $t : n \geq t > p$. Простым сложением показывается, что условная независимость (1) для элементарных событий $C = C_i$ и $D = D_j$ при всех i и j , где события C_i попарно несовместны между собой при разных i , события D_j попарно несовместны между собой при разных j , влечет условную независимость (1) для пары сумм событий $C = C_1 + C_2 + \dots + C_k$ и $D = D_1 + D_2 + \dots + D_l$. Таким образом, условная независимость C и D следует из условной независимости всех пар C_i и D_j . А в качестве этих событий мы можем взять самые маленькие события нужного нам вида.

Минимальное событие C_i , зависящее от всех $\xi_s : s \leq p$ состоит из всех траекторий ω , значения которых в промежутке $s \leq p$ составляют фиксированную траекторию ω_1 (т. е. все значения $\omega_1(s)$ зафиксированы при $s \leq p$).

Минимальное событие D_j , зависящее от всех $\xi_t : n \geq t > p$ состоит из всех траекторий, значения которых в промежутке $n \geq t > p$ составляют фиксированную траекторию ω_2 (т. е. все значения $\omega_2(t)$ зафиксированы при $n \geq t > p$).

Возможны две ситуации. В *первой ситуации* либо траектория ω_1 не кончается в точке (p, r) , либо траектория ω_2 не продолжается на момент времени p со значением r . В этих случаях события C_i или D_j несовместны с $A_{p,r}$ и тогда либо $\mathbb{P}(C_i|A_{p,r}) = 0$, либо $\mathbb{P}(D_j|A_{p,r}) = 0$ и равенство (1) очевидно, так как обе его части равны. 0.

Во *второй ситуации* обе траектории продолжаются на момент времени p со значением r . Поэтому $C_i \subset A_{p,r}$, $D_j \subset A_{p,r} \neq \emptyset$. Мы можем вычислить число траекторий в событиях $C_i A_{p,r}$ и $D_j A_{p,r}$ и сосчитать условные вероятности. Начнем с числа $|A_{p,r}|$. Число траекторий из точки $(0, 0)$ в (p, r) мы уже сосчитали, оно равно $C_p^{\frac{p+r}{2}}$. Но число способов продолжения каждой траектории на моменты времени от $p+1$ до n равно 2^{n-p} . Таким образом,

$$|A_{p,r}| = C_p^{\frac{p+r}{2}} 2^{n-p},$$

$$\mathbb{P}(A_{p,r}) = C_p^{\frac{p+r}{2}} 2^{n-p} / 2^n = C_p^{\frac{p+r}{2}} 2^{-p}.$$

Далее, $|C_i A_{p,r}|$ равно числу способов продолжения траектории ω_1 на $\overline{1, p}$ на все моменты времени $n \geq t > p$. Число таких продолжений равно 2^{n-p} . И, наконец, $|D_j A_{p,r}|$ равно числу таких продолжений траектории ω_2 на $s < p$, соединяющих $(0, 0)$ и (p, r) . Таким образом, $|D_j A_{p,r}| = C_p^{\frac{p+r}{2}}$. Резюмируем,

$$\mathbb{P}(C_i|A_{p,r}) = \frac{2^{n-p} 2^{-n}}{C_p^{\frac{p+r}{2}} 2^{-p}}, \quad \mathbb{P}(D_j|A_{p,r}) = \frac{C_p^{\frac{p+r}{2}} 2^{-n}}{C_p^{\frac{p+r}{2}} 2^{-p}},$$

$$\mathbb{P}(C_i D_j|A_{p,r}) = \frac{2^{-n}}{C_p^{\frac{p+r}{2}} 2^{-p}} = \mathbb{P}(C_i|A_{p,r}) \mathbb{P}(D_j|A_{p,r}).$$

Итак, (1) выполняется.

4. Задача о баллотировке.

В рамках схемы случайного блуждания мы решим следующую задачу:

Задача. У двух кандидатов, A и B , имеется ровно по n голосов. Эти голоса были поданы и происходит подсчет голосов. Какова вероятность события

$$A = \{\text{в течение подсчета } A \text{ будет всегда опережать } B\}.$$

Уточним, что в начальный момент времени и в последний момент времени у них одинаковое число голосов – по 0 и по n .

Элементарный исход $\omega \in \Omega$ в этой задаче — это траектория, которая в каждый момент времени показывает разность числа учтенных голосов за A и за B . Нарисуйте примерную траекторию.

Каждая такая траектория начинается в точке $(0, 0)$ и кончается в точке $(2n, 0)$. Траектории ω в интересующем нас событии A состоят из траекторий, которые всюду, за исключением 0 и $2n$, находятся выше оси x -ов. Итак,

$$|\Omega| = C_{2n}^n, \mathbb{P}(A) = \frac{|A|}{|\Omega|}, |A| = ?$$

Для подсчета $|A|$ мы используем принцип отражения. Очевидно, что каждый $\omega \in A$ проходит через точки $(1, 1)$ и $(2n - 1, 1)$. Легко считается число всех траекторий из $(1, 1)$ в $(2n - 1, 1)$ — во всех этих траекториях имеется $n - 1$ участков \searrow и $n - 1$ участков \nearrow . Поэтому число таких траекторий равно C_{2n-2}^{n-1} . Вычтем из него число траекторий из $(1, 1)$ в $(2n - 1, 1)$, по дороге касающихся оси x -ов, разность равна $|A|$. По принципу отражения

$$\begin{aligned} & \{\text{число траекторий из } (1, 1) \text{ в } (2n - 1, 1), \text{ по дороге касающихся оси } x\text{-ов}\} = \\ & = \{\text{число траекторий из } (1, 1) \text{ в } (2n - 1, -1)\} = C_{2n-2}^n. \end{aligned}$$

Итак, $|A| = C_{2n-2}^{n-1} - C_{2n-2}^n$.

Упражнение. Спички продаются по цене 50 копеек за коробку. $2n$ человек хотят купить по одной коробке, причем у n из них имеются 50-копеечные монеты, а у остальных n — монеты в 1 рубль. У продавца в данный момент нет ни одной 50-копеечной монеты. Какова вероятность того, что если наши $2n$ человек выстроются в очередь случайным образом, продавец сможет каждому из них в случае необходимости без задержки вернуть сдачу.

Указание. Очевидно, что первым в очереди должен оказаться человек с 50-копеечной монетой, а последним — с 1 рублем.

5. Задача о постоянном везении.

Два игрока играют совершенно одинаково и выигрывают друг у друга с вероятностью $1/2$. Сыграно $2n$ партий. Доказать, что вероятность того, что один из них будет всегда не позади другого равна вероятности того, что в $2n$ -й момент времени у обоих будет равное число выигрышей, то есть равна числу $C_n^n / 2^{2n}$.

Доказательство с помощью взаимно однозначного соответствия множества \tilde{A} всех траекторий не ниже оси x -в и множества A всех траекторий, кончающихся в точке $(2n, 0)$. Рассмотрим траекторию $\omega \in A$ и разделим ее на два участка: ω_1 — часть ω от начала до первого достижения инфимума, ω_2 — часть ω после первого достижения инфимума.

Часть ω_2 мы осторожно сдвигаем и переводим начало в точку $(0, 0)$, получаем часть $\tilde{\omega}_1$ траектории $\tilde{\omega}$. Часть ω_1 , которая заканчивается в точке $(r, -s)$, мы симметрично поворачиваем вокруг оси $x = r$, а после этого осторожно и параллельно сдвигаем с началом в конце куска траектории $\tilde{\omega}_1$. Получили кусок траектории $\tilde{\omega}_2$. Куски $\tilde{\omega}_1$ и $\tilde{\omega}_2$ составляют траекторию $\tilde{\omega} \in \tilde{A}$, которая сопоставляется траектории $\omega \in A$.

Обратное отображение строится в обратном порядке. Берем траекторию $\tilde{\omega} \in \tilde{A}$, которая кончается в точке $(2n, 2m)$ и лежит выше оси x -ов. Далее пересекаем ее прямой $\{y = m\}$, часть траектории $\tilde{\omega}$ до момента последнего по времени пересечения $\{y = m\}$ обозначим $\tilde{\omega}_1$,

часть после этого момента обозначим $\tilde{\omega}_2$. Часть $\tilde{\omega}_2$ поворачиваем и сдвигаем в начало, получаем часть ω_1 , к концу ω_1 сдвигаем $\tilde{\omega}_1$, получаем часть ω_2 . Вместе ω_1 и ω_2 составляют ω .

6. Лемма о возврате в начало.

Новые обозначения.

Введем событие:

$C_{2k,0} = \{\text{первое возвращение в начало } 0 \text{ произошло в момент } 2k\}$,

напомним, что

$A_{2k,0} = \{\text{возвращение траектории в начало } 0 \text{ в момент } 2k\}$.

Обозначим $f_{2k} = \mathbb{P}(C_{2k,0})$, $u_{2k} = \mathbb{P}(A_{2k,0})$, $f_0 = 0$, $u_0 = 1$.

Лемма. $u_{2n} = \sum_{k=0}^n f_{2k} u_{2n-2k}$.

Мы представим событие $A_{2n,0}$ в виде суммы несовместных событий. Очевидно, что

$$A_{2k,0} \supset C_{2k,0}, \quad A_{2n,0} \subset \sum_{k=0}^n C_{2k,0}.$$

Тогда

$$A_{2n,0} = \sum_{k=0}^n A_{2n,0} C_{2k,0} = \sum_{k=0}^n A_{2n,0} C_{2k,0} A_{2k,0}. \quad (*)$$

Здесь мы использовали $C_{2k,0} \subset A_{2k,0}$. Событие $A_{2n,0}$ относится к будущему процесса при настоящем времени $2k$, а событие $C_{2k,0}$ — к прошлому процесса. Мы вправе использовать утверждение об условной независимости прошлого и будущего (в нем $p = 2k$, $r = 0$). Используя это и формулу умножения (3.3) по разу в каждой строчке следующей формулы, имеем:

$$\begin{aligned} \mathbb{P}(A_{2n,0} C_{2k,0} A_{2k,0}) &= \mathbb{P}(A_{2n,0} C_{2k,0} | A_{2k,0}) \mathbb{P}(A_{2k,0}) = \\ &= \mathbb{P}(A_{2n,0} | A_{2k,0}) \mathbb{P}(C_{2k,0} | A_{2k,0}) \mathbb{P}(A_{2k,0}) = \mathbb{P}(A_{2n,0} | A_{2k,0}) \mathbb{P}(C_{2k,0}). \end{aligned}$$

Теперь проинтерпретируем полученное выражение. Второй множитель в правой части — это f_{2k} . А первый множитель — вероятность попадания из 0 в момент времени $2k$ в 0 в момент времени $2n$. Так как свойства процесса не меняются со временем, то эта вероятность совпадает с вероятностью попадания из 0 в момент времени 0 в 0 в момент времени $2n-2k$, то есть с числом u_{2n-2k} . (Последнее утверждение можно проверить и непосредственно, подсчитать $|A_{2n,0} A_{2k,0}| (= C_{2k}^k C_{2n-2k}^{n-k})$, разделить на $|A_{2k,0}| (= C_{2k}^k 2^{2n-2k})$ и разделить одно число на другое.) Далее используем $(*)$ и получаем

$$\mathbb{P}(A_{2n,0}) = \sum_{k=0}^n \mathbb{P}(A_{2n,0} C_{2k,0}) = \sum_{k=0}^n f_{2k} u_{2n-2k}.$$

Эта лемма будет очень существенной в решении задачи о возвращении в начало и в следующем пункте.

Упражнение. Проверьте перебором эту лемму для случая $n = 3$.

7. Задача о продолжительном везении. Распределение арксинуса.

Теперь мы вычислим вероятность того, что ровно на $2k$ участках траектория находилась выше оси x -в (где-то, может быть, ее касаясь), а на $2n - 2k$ участках траектория находилась ниже оси x -в. Обозначим:

$$\begin{aligned} p_{2k,2n} &= \mathbb{P}\{\text{ровно на } 2k \text{ участках траектория находилась выше оси } x\text{-в,} \\ &\quad \text{на } 2n - 2k \text{ участках траектория находилась ниже оси } x\text{-в}\}. \end{aligned}$$

Мы докажем следующее утверждение.

Теорема. $p_{2k,2n} = u_{2k} u_{2n-2k}$.

Лемма.

$$p_{2k,2n} = \frac{1}{2} \sum_{m=0}^k f_{2m} p_{2k-2m,2n-2m} + \frac{1}{2} \sum_{m=0}^{n-k} f_{2m} p_{2k,2n-2m}.$$

Доказательство повторяет рассуждения леммы 5. Мы делим наше событие на куски, в каждом из которых зафиксирован момент $2m$ первого касания оси x -в. При этом на первых $2m$ шагах траектория с равной вероятностью может быть выше оси x -в или ниже оси x -в, поэтому в каждая сумма умножается на $\frac{1}{2}$.

Из лемм 5 и 6 методом математической индукции выводится утверждение теоремы. Предположим, что для всех меньших n и всех $k \leq n$ утверждение теоремы верно. Тогда, используя лемму 6, мы получаем

$$\begin{aligned} p_{2k,2n} &= \frac{1}{2} \sum_{m=0}^k f_{2m} p_{2k-2m,2n-2m} + \frac{1}{2} \sum_{m=0}^{n-k} f_{2m} p_{2k,2n-2m} = \\ &= \frac{1}{2} \sum_{m=0}^k f_{2m} u_{2k-2m} u_{2n-2k} + \frac{1}{2} \sum_{m=0}^{n-k} f_{2m} u_{2k} u_{2n-2k-2m} = \\ &= \frac{1}{2} \left(\sum_{m=0}^k f_{2m} u_{2k-2m} \right) u_{2n-2k} + \frac{1}{2} \left(\sum_{m=0}^{n-k} f_{2m} u_{2n-2k-2m} \right) u_{2k} = \\ &= \frac{1}{2} u_{2k} u_{2n-2k} + \frac{1}{2} u_{2n-2k} u_{2k}, \end{aligned}$$

что и требовалось. Заметим, что, используя индукцию, мы неявно используем и утверждение пункта 5. Оно доказывает справедливость нашего утверждения при всех n и $k = n$. Благодаря этому, вычисляется вероятность отсутствия пересечения оси x -ов после момента последнего пересечения.

Теперь мы проанализируем полученный результат и рассмотрим, что получается в пределе.

Предложение. $p_{2k,2n} \approx \frac{1}{\pi \sqrt{k(n-k)}}$.

Мы используем формулу Стирлинга:

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e} \right)^n,$$

Поэтому

$$\begin{aligned} u_{2k} &= C_{2k}^k 2^{-2k} \approx \frac{\left(\frac{2k}{e}\right)^{2k} \sqrt{2\pi 2k}}{\left(\frac{k}{e}\right)^k \sqrt{2\pi k} \left(\frac{k}{e}\right)^k \sqrt{2\pi k}} 2^{-2k} = \frac{1}{\sqrt{\pi k}}, \\ p_{2k,2n} &\approx \frac{1}{\pi \sqrt{k(n-k)}}. \end{aligned}$$

Случайная величина, которую мы рассмотрели, принимает значения $2k$. Давайте переведем эту величину в отрезок $[0,1]$, заменив $2k$ на $x_k = 2k/2n$. Тогда

$$\mathbb{P}_n\{x_k\} \approx \frac{1}{\pi \sqrt{x_k(1-x_k)}} (1/n).$$

Теперь вычислим

$$\int_0^1 f(x) d\mathbb{P}_n(x) \approx \sum_{k=1}^n f(x_k) \frac{1}{\pi \sqrt{x_k(1-x_k)}} (x_{k+1} - x_k).$$

Это очень напоминает интегральные суммы Римана для интеграла

$$\int_0^1 f(x) \frac{1}{\pi \sqrt{x(1-x)}} dx$$

(хотя интеграл является несобственным и это не совсем корректно). Здесь (по определению слабой сходимости) f — непрерывная ограниченная функция на отрезке $[0, 1]$. Мы этого доказывать не будем, но при $n \rightarrow \infty$ вероятность \mathbb{P}_n будет слабо сходиться к распределению с плотностью $\frac{1}{\pi \sqrt{x(1-x)}}$ на интервале $(0, 1)$. Функция распределения будет выражаться через арксинус, поэтому полученное распределение носит название 'закон арксинуса'.

8. Задача о возвращении в начало.

Теорема. *Пьяный гуляка возвращается в начало с вероятностью 1 тогда и только тогда, когда $\sum_n u_{2n} = \infty$.*

Мы введем две функции, связанные со случайным блужданием, которые называются производящими функциями.

Но прежде всего уточним постановку задачи. Вероятность возврата в начало за какое-то число шагов мы можем представить как сумму вероятностей первого возвращения за 2, 4, ... шагов. Эти вероятности мы обозначали f_{2n} . Заметим, что в этой задаче у нас нет одной на все случаи модели. Число шагов может стремиться к ∞ , при увеличении числа шагов увеличивается число элементарных исходов в вероятностной модели.

Итак, нужно выяснить, когда $\sum_{2n} f_{2n} = 1$. Введем две функции: F и U , заданные на $(0, 1)$.

$$F(z) = f_0 + f_2 z^2 + f_4 z^4 + \dots, \quad U(z) = u_0 + u_2 z^2 + u_4 z^4 + \dots$$

Замечание. Введенные функции похожи на производящие функции, которые часто используемые при изучении распределений случайных величин, равных $0, 1, 2, \dots, n, \dots$ с вероятностями p_n , где $\sum_n p_n = 1$. Продолжающая функция задается формулой $\psi(z) = \sum_n p_n z^n$ и используется вместо характеристической функции. Числа p_n совпадают со значениями n -х производных в нуле функции $\psi(z)$. В нашем случае сумма u_n заведомо не равна 1, так как эти числа являются вероятностями событий, не являющихся несовместными. А для чисел f_n равенство суммы 1 является вопросом, ответ на который мы пытаемся получить.

Из леммы 5 непосредственно выводится сравнением коэффициентов при степенях z равенство

$$F(z)U(z) = U(z) - 1.$$

Поэтому

$$U(z) = \frac{1}{1 - F(z)}.$$

Устремим в этом равенстве $z \rightarrow 1$. Имеем:

$$u_0 + u_2 + u_4 + \dots = \frac{1}{1 - (f_0 + f_2 + f_4 + \dots)}.$$

Анализируя это равенство, мы получаем теорему: ряд в левой части расходится тогда и только тогда, когда в правой части знаменатель сходится к 0.

9. Следствия.

Из расходимости

$$\sum_n u_{2n} = \sum_n \frac{1}{\sqrt{\pi n}} = \infty$$

следует, что в симметричном случайному блужданию размерности 1 пьяный гуляка с вероятностью 1 вернется домой.

В размерности 2 пьяный гуляка в каждый момент времени переходит из точки (x, y) в точки $(x + 1, y)$, $(x - 1, y)$, $(x, y + 2)$, $(x, y - 1)$, причем в каждую из них с вероятностью $1/4$. Анализ этого случая удобно провести сведением его к одномерному. Оказывается, двумерное движение пьяного гуляки можно представить как два независимых одномерных движения. Речь не идет о движении по координатам (движения по координатам x и y зависят, двигаясь по одной, мы не двигаемся по другой координате). Но мы можем рассмотреть два движения по прямым $y = x$ и $y = -x$. По каждой из них в любой момент времени мы двигаемся с вероятностью $1/2$ в одну или другую сторону на расстояние $1/\sqrt{2}$. В результате мы получим движение пьяного гуляки на плоскости.

Чтобы вернуться в начало за $2n$ шагов, пьяный гуляка должен вернуться в начало по каждой из прямых $x = y$ и $x = -y$. Таким образом, вероятность возвращения равна

$$u_{2n} \approx \frac{1}{\pi n}.$$

Ряд из этих вероятностей по прежнему расходится, по теореме 8 пьяный гуляка на плоскости с вероятностью 1 возвращается в начало.

Случай размерности 3 не сводится к одномерному и дает другой ответ. В трехмерном пространстве на каждом шаге пьяный гуляка имеет шесть вариантов: вверх или вниз по координате x , то же – по y , то же – по z . Вероятность каждого направления мы считаем равной $1/6$. За $2n$ шагов общее число возможных траекторий равно 6^{2n} , вероятность каждой траектории равна 6^{-2n} .

Легко видеть, что в событие {возврат в начало за $2n$ шагов} входят элементарные исходы, содержащие i шагов по оси x вверх, i шагов по оси x вниз, j шагов по оси y вверх, j шагов по оси y вниз, k шагов по оси z вверх, k шагов по оси z вниз, причем должно быть $i + j + k = n$. Число таких исходов было подсчитано в полиномиальной модели (шесть клеток, $2n$ кроликов).

Итак, наша вероятность u_{2n} возврата в начало ровно за $2n$ шагов равна:

$$u_{2n} = 6^{-2n} \sum_{i+j+k=n} \frac{2n!}{i!j!k!} = 6^{-2n} \sum_{i+j+k=n} \frac{2n!}{n!n!} \left(\frac{n!}{i!j!k!} \right)^2 \quad (2)$$

Если мы один из множителей в квадрате заменим на его максимальное возможное значение, все выражение разве что увеличится. Имеем:

$$u_{2n} \leq 6^{-2n} \frac{2n!}{n!n!} \max \left(\frac{n!}{i!j!k!} \right) \sum_{i+j+k=n} \left(\frac{n!}{i!j!k!} \right) = 2^{-2n} 3^{-n} \frac{2n!}{n!n!} \max \left(\frac{n!}{i!j!k!} \right) = O \left(\frac{1}{n^{3/2}} \right).$$

Здесь мы использовали формулу Стирлинга

$$n! \approx \left(\frac{n}{e} \right)^n \sqrt{2\pi n}.$$

Но предварительно полезно заметить, что максимум достигается, когда все три слагаемых i , j , k примерно равны между собой, то есть примерно равны числу $n/3$. Ряд из полученных оценок сходится.

Упражнение. Проверьте, что если $i < n/3$, $j > n/3$, то $i!j!k! < (i-1)!(j+1)!k!$. (Разумеется, $i + j + k = n$, $i, j, k \in \mathbf{N}$.

§29. Цепи Маркова

Марковский случайный процесс — это процесс, для которого любое событие, записывающееся через значения процесса в будущем, при условии фиксации значения процесса в настоящий момент не зависит от события, записывающегося через значения процесса в прошлом.

Мы будем использовать для такого процесса $\eta(\cdot)$ следующие обозначения:

$p(s, x; t, y)$ — условная плотность случайной величины η_t в точке y при условии, что $\eta_s = x$,

$$P(s, i; t, j) = \mathbb{P}\{\eta(t) = j | \eta(s) = i\}$$

1. Равенство Колмогорова — Чепмена.

Пусть $s < u < t$. Тогда

$$p(s, x; t, y) = \int_{-\infty}^{\infty} p(s, x; u, z)p(u, z; t, y)dz.$$

В непрерывном случае для доказательства уравнения Колмогорова — Чепмена нужно приложить определенные усилия. В дискретном случае счетного числа состояний все очень просто и мы докажем дискретный аналог этого равенства:

$$P(s, i; t, j) = \sum_k P(s, i; u, k)P(u, k; t, j) \quad (1).$$

Но для этого нам придется вернуться к обозначениям в виде условных вероятностей, для которых (1) выглядит следующим образом:

$$\mathbb{P}\{\eta(t) = j | \eta(s) = i\} = \sum_k \mathbb{P}\{\eta(u) = k | \eta(s) = i\} \mathbb{P}\{\eta(t) = j | \eta(u) = k\} \quad (2).$$

Это равенство напоминает формулу полной вероятности, но отличается от нее тем, что все вероятности являются условными, причем с разными условиями. Давайте запишем вместо (2) формулу полной вероятности, но не по первоначальной вероятности \mathbb{P} , а по условной вероятности $\mathbb{P}\{\cdot | \eta(s) = i\}$. Имеем:

$$\begin{aligned} &\mathbb{P}\{\eta(t) = j | \eta(s) = i\} = \\ &= \sum_k \mathbb{P}\{\eta(t) = j | \eta(u) = k, \eta(s) = i\} \mathbb{P}\{\eta(u) = k | \eta(s) = i\}. \end{aligned}$$

В этом равенстве единственная разница с нужным нам (2) состоит в том, что вместо $\mathbb{P}\{\eta(t) = j | \eta(u) = k\}$ написано $\mathbb{P}\{\eta(t) = j | \eta(u) = k, \eta(s) = i\}$. Но по нашему определению марковского процесса вероятность будущего $\{\eta(t) = j\}$ при фиксированном настоящем $\{\eta(u) = k\}$ не зависит от прошлого $\{\eta(s) = i\}$. Поэтому

$$\mathbb{P}\{\eta(t) = j | \eta(u) = k, \eta(s) = i\} = \mathbb{P}\{\eta(t) = j | \eta(u) = k\}.$$

Кстати, для марковского процесса с совместными плотностями марковость выглядит следующим образом: при $s_1 < s_2 < \dots < s_n < u < t_1 < t_2 < \dots < t_m$

$$p_{\eta(t_1) \dots \eta(t_m)}(y_1, \dots, y_m | \eta(s_1) = x_1, \dots, \eta(s_n) = x_n, \eta(u) = z) = p_{\eta(t_1) \dots \eta(t_m)}(y_1, \dots, y_m | \eta(u) = z).$$

2. Однородная марковская цепь с дискретным временем.

Мы вернемся к более простой ситуации дискретного времени $\{0\} \cup \mathbf{N}$ и дискретного множества состояний \mathbf{N} . Последнее носит довольно условный характер, числа выступают в виде знаков и нумерация состояний не зависит от их свойств. Это не всегда удобно, например, при изучении одномерного случайного блуждания состояния были целыми числами — элементами \mathbf{Z} , а в двумерном случае — элементами \mathbf{Z}^2 . Мы будем рассматривать самый простой вариант и считать, что свойства системы не меняются со временем. Посмотрим, какие задачи при этом возникают.

Следующие обозначения совершенно естественны:

Изучаемый случайный процесс обозначим ξ_n , $n = 0, 1, 2, \dots$ и будем его называть *однородной цепью Маркова со счетным числом состояний*.

p_{ij} — вероятность перехода системы за один шаг из состояния i в состояние j .

Другими словами, $p_{ij} = \mathbb{P}(\xi_{n+1} = j | \xi_n = i)$.

$p_{ij}(n)$ — вероятность перехода системы за n шагов из состояния i в состояние j , в таком случае $p_{ij} = p_{ij}(1)$. Очевидно, что должно выполняться $\sum_j p_{ij} = 1$ для любого i , так как из состояния i система обязательно куда-нибудь перейдет.

Как мы уже установили выше, для цепей Маркова справедливо равенство Чепмена — Колмогорова:

$$p_{ij}(m+n) = \sum_k p_{ik}(m)p_{kj}(n).$$

Матрицу $[p_{ij}(n)]$ переходных вероятностей за n шагов мы обозначим через $\mathcal{P}(n)$, $\mathcal{P} = \mathcal{P}(1)$.

Из равенства Чепмена — Колмогорова следует (по индукции):

$$\mathcal{P}(n) = \mathcal{P}^n.$$

Действительно, в равенстве Чепмена — Колмогорова мы используем операцию произведения матриц.

Важнейшим примером однородной цепи Маркова является одномерное случайное блуждание, в этом случае множество состояний обозначается \mathbf{Z} , а переходные вероятности имеют вид: $p_{i(i+1)} = p_{i(i-1)} = 1/2$. Интересным вопросом теории однородных цепей Маркова является вопрос о возвратных и невозвратных состояниях. Мы его по-существу рассмотрели при изучении случайного блуждания. Вывод критерия использует лишь свойства цепей Маркова. Итак, состояние i_0 является *возвратным* (то есть вероятность возврата в него (хотя бы раз) за неограниченное время равна 1) тогда и только тогда, когда

$$\sum_n p_{i_0 i_0}(n) = \infty.$$

Другой важный вопрос — о существовании стационарных вероятностей на множестве состояний. Дело в том, что стандартная ситуация при изучении цепей Маркова — наличие лишь переходных вероятностей. Но при этом не определены вероятности самих состояний в любой момент времени n , то есть вероятностей

$$p_i(n) = \mathbb{P}\{\xi_n = i\}.$$

Очевидно, что нам достаточно знать эти вероятности лишь при $n = 0$. Будем обозначить $p_i = p_i(0)$.

Тогда по формуле полной вероятности (без использования марковости) мы имеем:

$$p_j(n) = \sum_i p_i p_{ij}(n), \quad (7)$$

а вектор вероятностей $[p_j(n)]$ вычисляется по формуле

$$[p_i(n)] = [p_i] \mathcal{P}^n.$$

Упражнение 2. Марковская цепь имеет три состояния $\{1, 2, 3\}$, вектор начальных вероятностей имеет вид $[1/2, 1/3, 1/6]$, матрица переходных вероятностей имеет вид

$$\mathcal{P} = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

Используя определение цепи Маркова, вычислить $\mathbb{P}\{\xi_1 = 1, \xi_2 = 2, \xi_3 = 3\}$.

3. Существование стационарного распределения.

Вектор вероятностей $[p_i]$ называется *начальным распределением*. Напомним, что начальное распределение для случайного блуждания сосредоточено в точке 0. Может так получиться, что для некоторого начального распределения $[p_i]$ распределения в другие моменты времени, вычисляемые по формуле (7), не меняются:

$$p_i(n) = p_i.$$

Такое начальное распределение называется *стационарным*.

Имеется несколько интересных вопросов:

- 1) Существуют ли стационарные распределения?
- 2) Если существуют, то сколько их? Точнее, какова размерность множества стационарных распределений? Значения скольких параметров однозначно задают распределение?

Приведем сначала примеры. Можно показать, что у процесса случайного блуждания стационарных распределений быть не может. Доказательство можно провести от противного: так как $p_i \rightarrow 0$, когда $i \rightarrow \infty$ или $i \rightarrow -\infty$, то существует максимальное p_{i_0} , причем можно считать, что оно крайнее справа среди максимальных, то есть $p_{i_0-1} \leq p_{i_0} > p_{i_0+1}$. Тогда на следующем шаге

$$p_{i_0}(1) = \frac{p_{i_0-1} + p_{i_0+1}}{2} < p_{i_0}(0).$$

В то же время для систем с конечным числом состояний легко построить пример существования стационарного распределения. Пусть, например, система содержит два состояния, $\{0, 1\}$,

$$p_{0,1} = 1, p_{1,0} = 1.$$

Тогда стационарным распределением является распределение

$$p_0 = p_1 = 1/2.$$

Имеется знаменитая теорема Брауэра, которая утверждает, что *любое непрерывное отображение конечномерного выпуклого компакта в себя имеет хотя бы одну неподвижную точку*. Множество всех вероятностей на конечном (n) числе состояний очевидно выпукло и компактно. Более того, это множество является симплексом с n вершинами $(1, 0, \dots, 0)$, $(0, 1, \dots, 0)$, $(0, 0, \dots, 1)$.

Таким образом, цепь Маркова с конечным числом состояний всегда имеет хотя бы одно стационарное распределение.

Но можно привести нехитрый пример, когда число стационарных распределений бесконечно:

$$\{0, 1, 2, 3\}, p_{0,1} = 1, p_{1,0} = 1, p_{2,3} = 1, p_{3,2} = 1.$$

Стационарные распределения —

$$p'_0 = p'_1 = 1/2; p''_2 = p''_3 = 1/2.$$

Другие стационарные распределения являются выпуклыми комбинациями данных двух.

В некоторых ситуациях (но не всегда, контрпримеры вы можете найти выше) любое начальное распределение с ростом времени начинает сходиться к некоторому стационарному распределению. Представляет интерес о скорости сходимости и о том, начиная с какого времени мы можем считать распределение приблизительно стационарным.

Интересный пример такой цепи представляет собой мешание колоды из 32 карт. За один шаг мы берем случайнм образом делим колоду на три части — A, B, C и переставляем части в обратном порядке — C, B, A . Ясно, что таким способом мы можем получить не любые перестановки карт, т. е., нет перестановок, вероятности перехода от любой другой перестановки карт не нуль.

Множество состояний состоит из всех перестановок, то есть из $32!$ элементов. На каждом шаге система из данного состояния может перейти примерно (точно считать долго) в $c = C_{32}^2$ состояния, причем вероятности перехода мы считаем равными $1/c$. Начальная вероятность сидит в одной фиксированной перестановке. Но при $n \rightarrow \infty$ вероятность сходится к стационарному распределению, для которого вероятность любой перестановки равна $1/32!$. Представляется совершенно очевидным, что такое стационарное распределение в данной ситуации является единственным. Но из ниже приведенной теоремы немедленно это не следует, так как в данном случае условия теоремы не выполняются. В качестве упражнения можно проверить

существование достаточно большой итерации оператора \mathcal{P} , которая удовлетворяет условиям теоремы с очень маленьким λ и для любого из $32!$ состояний. Отсюда следует существование единственного инвариантного стационарного распределения для \mathcal{P} . (Это следует из равенства $\mathcal{P}^n\mathcal{P} = \mathcal{P}\mathcal{P}^n$.) В этой задаче можно оценить коэффициент λ , очевидно, что он очень мал, поэтому скорость сходимости к стационарному распределению очень мала. Надо очень долго мешать карты.

В конце этого краткого экскурса я приведу (с двумя доказательствами, с выкладками, другое, использующее идеи функционального анализа, но без выкладок) одно утверждение о существовании и единственности стационарного распределения.

Теорема. *Допустим, что существует особое состояние i_0 со свойством:*

$$\inf_i p_{i,i_0} \geq \lambda > 0.$$

Тогда стационарное распределение существует и оно единствено.

Первое доказательство состоит в применении принципа Банаха сжатых отображений к преобразованию

$$P[p_i] := [p_i]\mathcal{P}, \quad (8)$$

пространства начальных вероятностей. Прежде чем применять принцип Банаха, надо ввести в пространстве всех начальных вероятностей метрику и проверить полноту этого пространства в данной метрике. Введем метрику следующим образом:

$$d(p', p'') = \sum_i |p'_i - p''_i|.$$

Проверка аксиом метрики и полноты соответствующего пространства вероятностей очевидна. Теперь оценим

$$d(p'\mathcal{P}, p''\mathcal{P}) = \sum_j \left| \sum_i (p'_i - p''_i) p_{ij} \right| = \sum_{j \neq i_0} \left| \sum_i (p'_i - p''_i) p_{ij} \right| + \left| \sum_i (p'_i - p''_i) p_{i_0 i_0} \right|.$$

Мы будем далее использовать, что

$$\sum_i p'_i = \sum_i p''_i = 1, \sum_j p_{ij} = 1 \text{ для всех } i.$$

Займемся второй суммой:

$$\begin{aligned} \left| \sum_i (p'_i - p''_i) p_{i_0 i_0} \right| &= \left| \sum_i (p'_i - p''_i) (p_{i_0 i_0} - \lambda + \lambda) \right| \leq \left| \sum_i (p'_i - p''_i) (p_{i_0 i_0} - \lambda) \right| + \left| \sum_i (p'_i - p''_i) \lambda \right| = \\ &= \left| \sum_i (p'_i - p''_i) (p_{i_0 i_0} - \lambda) \right| + 0 \leq d(p', p'')(p_{i_0 i_0} - \lambda) = d(p', p'')p_{i_0 i_0} - d(p', p'')\lambda. \end{aligned}$$

Теперь оцениваем все:

$$d(p'\mathcal{P}, p''\mathcal{P}) \leq (1 - \lambda)d(p', p'').$$

Мы воспользовались очевидной оценкой

$$\sum_j \left| \sum_i (p'_i - p''_i) p_{ij} \right| \leq \sum_j \sum_i |p'_i - p''_i| p_{ij} = \sum_i \sum_j |p'_i - p''_i| p_{ij} = \sum_i |p'_i - p''_i|.$$

Второе доказательство также состоит в применении принципа Банаха, но для оценок мы используем экстремальные точки единичного шара в одном банаховом пространстве и оцениваем норму оператора \mathcal{P} . В случае счетного множества состояний мы можем считать вероятности элементами пространства l_1 , и

$$d(p', p'') = \|p' - p''\|_{l_1}.$$

Это множество является пересечением конуса неотрицательных элементов $\{(x_i) : x_1 \geq 0, x_2 \geq 0, \dots\}$ со сдвинутым линейным подпространством $\{(x_i) : \sum_n x_n = 1\}$.

Упражнение. Проверьте, что и конус и сдвинутое линейное подпространство замкнуты в l_1 .

Итак, множество вероятностей — замкнутое подмножество полного пространства l_1 , поэтому само является полным. Очевидно, что преобразование P является линейным оператором на l_1 , а для любого линейного оператора T , действующего в банаховом пространстве B с нормой $\|\cdot\|$ и задаваемой ею метрикой $d(x, y) := \|x - y\|$ имеет место неравенство:

$$d(Tx, Ty) = \|T(x) - T(y)\| = \|T(x - y)\| \leq \|T\| \|x - y\| = \|T\| d(x, y).$$

Итак, казалось бы достаточно вычислить норму линейного оператора P в пространстве l_1 и доказать, что $\|P\| < 1$. Но все не так просто. Во первых, любой линейный оператор имеет одну неподвижную точку — 0, а по принципу Банаха, если его норма меньше 1, у него не может быть других неподвижных точек, в частности, вероятности не могут быть неподвижными точками. Впрочем, норма оператора P на самом деле равна 1, в чем нетрудно убедиться. Немедленно проверяется, что $\|p\| \geq 1$, так как

$$P(1, 0, \dots) = [p_{1i}]_{i=1}^{\infty}, \sum_i p_{1i} = 1.$$

Таким образом, наша идея оказывается несостоятельной. Давайте уточним эту идею — на самом деле нас интересуют лишь значения $\|P(p' - p'')\|$, то есть норма значений оператора не на всем l_1 . Нетрудно видеть, что $\sum_i (p'_i - p''_i) = 0$, таким образом, нам нужна норма P на подпространстве l_1 — пространстве $l_1 \cap \{(x_i) : \sum_i x_i = 0\}$. Это подпространство не содержит точку $(1, 0, \dots)$, поэтому мы не потеряли шанс на то, что норма сужения P на это подпространство окажется строго меньше 1.

Теперь зайдемся вычислением нормы. Хорошо известно, что для вычисления нормы линейного оператора T достаточно рассмотреть сужение оператора на единичную сферу банахова пространства:

$$\sup_{\|x\| \leq 1} \|T(x)\| = \sup_{\|x\|=1} \|T(x)\| = \|T\|. \quad (*)$$

Более того, так как норма — выпуклая функция

$$\|\lambda x + \mu y\| \leq \lambda \|x\| + \mu \|y\|, \lambda + \mu = 1, \lambda \geq 0, \mu \geq 0.$$

Поэтому (это важно, если $\|x\| = \|y\| = 1$)

$$\|T(\lambda x + \mu y)\| \leq \lambda \|T(x)\| + \mu \|T(y)\| \leq \max\{\|T(x)\|, \|T(y)\|\}, \lambda + \mu = 1, \lambda \geq 0, \mu \geq 0.$$

Методом математической индукции это неравенство обобщается также на конечные и даже счетные выпуклые комбинации. Таким образом, если у нас имеется счетное или несчетное множество точек на единичной сфере, которые непредставимы в виде $\lambda x + \mu y$, где $\|x\| = \|y\| = 1$, $\lambda + \mu = 1$, $\lambda \geq 0$, $\mu \geq 0$ (такие точки мы будем называть *экстремальными*), и все точки единичного шара представляются как пределы выпуклых комбинаций экстремальных точек, то супремум в определении нормы линейного оператора достаточно взять лишь по множеству всех экстремальных точек. Нетрудно видеть, что в единичном шаре пространства l_1 экстремальными точками являются лишь точки вида $\pm e_n$, $e_n = (0, \dots, 0, 1, 0, \dots)$ (1 на n -м месте, остальные — нули). Легко проверяется, что все остальные точки единичного шара являются бесконечными выпуклыми комбинациями экстремальных точек. В единичном шаре пространства $l_1 \cap \{(x_i) : \sum_i x_i = 0\}$ экстремальными точками являются точки вида $\frac{1}{2}(e_m - e_n)$, $m \neq n$. Итак, достаточно оценить

$$\left\| P\left(\frac{1}{2}(e_m - e_n)\right) \right\| = \left\| \frac{1}{2} \left(\sum_i p_{mi} e_i - \sum_i p_{ni} e_i \right) \right\| =$$

$$\begin{aligned}
&= \frac{1}{2} \sum_i |p_{mi} - p_{ni}| = \frac{1}{2} \left(\sum_{i \neq i_0} + \sum_{i=i_0} \right) \leq \\
&\leq \frac{1}{2} \left(\sum_{i \neq i_0} (p_{mi} + p_{ni}) + \sum_{i=i_0} |p_{mi} - p_{ni}| \right) \leq \frac{1}{2} \left(\sum_{i \neq i_0} (p_{mi} + p_{ni}) + \sum_{i=i_0} (p_{mi} + p_{ni} - 2\lambda) \right) = \\
&= \frac{1}{2} \left(\sum_i (p_{mi} + p_{ni}) - 2\lambda \right) = 1 - \lambda.
\end{aligned}$$

Итак, норма оператора P на пространстве $l_1 \cap \{(x_i) : \sum_i x_i = 0\}$ не превосходит числа $1 - \lambda$,

$$d(p', p'') = \|P(p' - p'')\| \leq \|P\| d(p', p'') \leq (1 - \lambda) d(p', p'').$$

Это доказывает теорему.

Замечание. В примере с 32 картами условия доказанной теоремы не выполняются. Действительно, при регулярном мешании карт из последовательности $\{a, b, c\}$, где каждая из букв — некоторый набор карт в определенном порядке, мы получаем последовательность $\{c, b, a\}$. Число вариантов, которые можно получить за одно перемешивание, можно грубо оценить сверху — оно не превосходит 32^2 . Вероятности перехода в другие состояния равны 0. А общее число состояний равно $32!$. Однако некоторая большая итерация преобразования \mathcal{P} удовлетворяет условиям теоремы, а следовательно, является оператором сжатия. Итак, существует единственное стационарное распределение. (Нетрудно проверить, что за 32 стандартные операции мешания мы можем получить любой порядок карт. Таким образом, из начального состояния в любое другое мы переходим с вероятностью, большей чем $\lambda = 32^{32^2}$.) Заметим, что из теоремы Брауэра единственность не следует.

§30. Два замечательных процесса с непрерывным временем

После процесса случайного блуждания мы рассмотрим и найдем распределения двух важнейших процессов с непрерывным временем.

Эти процессы, пока мы их обозначаем η_t (случайная величина — значение процесса в момент времени t) очень сильно отличаются друг от друга, тем не менее, два первых условия на оба процесса совпадают:

- 1) независимость приращений: для любых моментов времени $t_1 < t'_1 \leq t_2 < t'_2 \leq \dots$ случайные величины

$$\eta_{t'_1} - \eta_{t_1}, \eta_{t'_2} - \eta_{t_2}, \dots$$

независимы. В частности, среднее произведения указанных случайных величин равно произведению средних, совместная плотность этих величин равна произведению плотностей, дисперсия суммы этих величин равна сумме дисперсий, характеристическая функция суммы этих величин равна произведению характеристических функций.

- 2) стационарность распределений приращений (они не зависят от времени): распределения случайных величин $\eta_{t+h} - \eta_t$ не зависят от t . В частности,

$$\mathbb{P}_{\eta_{t+h} - \eta_t} = \mathbb{P}_{\eta_h}$$

для любых неотрицательных t и h . Так как среднее и дисперсия определяются только распределениями,

$$\mathbb{E}(\eta_{t+h} - \eta_t) = \mathbb{E}\eta_h, \mathbb{D}(\eta_{t+h} - \eta_t) = \mathbb{D}\eta_h$$

для любых t и h .

Другие свойства двух процессов различаются, чтобы их сформулировать, мы должны рассказать о явлениях, которые эти процессы описывают.

1. Стандартный пуассоновский процесс можно представить как число η_t поломок станка за время $(0, t]$. Но у этого станка имеются свойства 1) и 2), которые выполняются, если станок является ткацким.

1) Приращения $\eta_{t'} - \eta_t$ — число поломок станка за время $(t, t']$ независимы. У реальных приборов это не так, если прибор с браком или почти с браком (разрешенные допуски были почти достигнуты), то прибор будет ломаться чаще, большое число поломок в начале работы будет означать, что и дальше он должен ломаться чаще. Для ткацкого станка это не так, поломки (обрывы нити) связаны со свойствами нити, а не станка.

2) означает, что станок не стареет, это условие будет разумно, если мы наблюдаем станок не слишком продолжительное время.

Замечание. Мы можем использовать эту модель для других задач, например, числа покупок, телефонных звонков и т. д., но в зависимости от времени дня или дня недели нужно менять масштаб времени, где-то оно должно течь медленно, а где-то — быстро.

3) аксиома процесса η_t также связана с интерпретацией процесса. Значения η_t являются целыми неотрицательными числами, поэтому имеют смысл вероятности $\mathbb{P}\{\eta_t = k\}$. Мы предполагаем, что за малое время t :

$$\mathbb{P}\{\eta_t = 1\} = \lambda t + o(t),$$

в то же время

$$\mathbb{P}\{\eta_t > 1\} = o(t).$$

Теперь мы составим бесконечную систему дифференциальных уравнений для всех функций

$$p_k(t) = \mathbb{P}\{\eta_t = k\}.$$

Рассмотрим

$$\begin{aligned} p_k(t + \Delta t) &= \mathbb{P}\{\eta_{t+\Delta t} = k\} = \mathbb{P}\{\eta_t = k, \eta_{t+\Delta t} - \eta_t = 0\} + \\ &+ \mathbb{P}\{\eta_t = k - 1, \eta_{t+\Delta t} - \eta_t = 1\} + \mathbb{P}\{\eta_t = k - 2, \eta_{t+\Delta t} - \eta_t = 2\} + \dots \end{aligned}$$

Далее мы используем аксиомы 1) и 2) для представления

$$\mathbb{P}\{\eta_t = k - i, \eta_{t+\Delta t} - \eta_t = i\} = \mathbb{P}\{\eta_t = k - i\} \mathbb{P}\{\eta_{\Delta t} = i\}$$

и 3) для представления $\mathbb{P}\{\eta_{\Delta t} = 0\}, \mathbb{P}\{\eta_{\Delta t} = 1\}, \mathbb{P}\{\eta_{\Delta t} > 1\}$. Получаем:

$$p_k(t + \Delta t) = p_k(t)(1 - \lambda \Delta t + o(\Delta t)) + p_{k-1}(t)(\lambda \Delta t + o(\Delta t)) + o(\Delta t).$$

Составляем дифференциальное уравнение:

$$\frac{dp_k(t)}{dt} = -\lambda p_k(t) + \lambda p_{k-1}(t).$$

При $k = 0$ уравнение решается разделением переменных,

$$p_0(t) = C e^{-\lambda t},$$

подстановка $t = 0$ дает $C = 1$. Для других k уравнения решаются методом вариации постоянной, при этом используется индуктивный процесс. Опираясь на знание $p_k(\cdot)$, мы находим $p_{k+1}(\cdot)$

$$p_k(t) = (\lambda t)^k e^{-\lambda t} + C_1,$$

подстановка $t = 0$ дает $C_1 = 0$.

Итак, η_t имеет распределение Пуассона с параметром $\lambda = \lambda t$.

2. Процесс Винера является частным случаем (при некоторых значениях параметров) процесса одномерного броуновского движения. Норберт Винер внес большой вклад в математическую теорию броуновского движения, в частности, доказал непрерывность траекторий

процесса броуновского движения. Итак, значением процесса (мы будем его обозначать η_t) является значение координаты x броуновской частицы в момент времени t , причем значение координаты в момент времени 0 мы будем считать 0. Броуновские частицы двигаются в однородной жидкости, свойства которой не зависят от времени и от координаты (однородность). Кроме того, мы считаем, что число частиц бесконечно, и поэтому в любой, даже самый маленький, промежуток времени о частицу ударяется бесконечное число частиц жидкости.

Все это лишь наводящие соображения, которые мы должны превратить в аксиомы процесса η_t на языке теории вероятностей. Итак, что означает независимость свойств жидкости от времени. Разумеется, это не означает равенства $\eta_t = \eta_s \forall t, s$ даже с вероятностью 1 (тогда броуновская частица должна оставаться на месте). Передвижение частицы за промежуток времени $[t, t+h]$ равно $\eta_{t+h} - \eta_t$, но независимость от времени не должна означать также равенства $\eta_{t+h} - \eta_t = \eta_{s+h} - \eta_s \forall t, s, h$ (это означало бы движение броуновской частицы с постоянной скоростью). Независимость свойств жидкости от времени дает нам аксиому 2) (см. п. 1): при данном $h \geq 0$ распределение приращения $\mathbb{P}_{\eta_{t+h}-\eta_t}$ одно и то же для всех $t \geq 0$.

Мы будем предполагать, что случайные величины η_t также удовлетворяют аксиоме 1), (независимость приращений процесса в непересекающиеся промежутки времени $[t_1, t'_1]$ и $[t_2, t'_2]$), которая означает отсутствие у броуновских частиц какой-либо инерционности, то есть скорости. Согласно этой аксиоме путь, пройденный частицей за промежуток времени $[t_1, t'_1]$ никак не влияет на путь за промежуток $[t_2, t'_2]$ (например, средняя скорость частицы может оказаться большой в одном промежутке и маленькой в другом), Это возможно лишь при наличии бесконечного числа соударений броуновской частицы с другими частицами за конечное время, которые и определяют траекторию частицы. На деле, молекул в воде конечное число, но чрезвычайно большое. Таким образом, аксиома 1) вполне разумна. Аксиома 2) отнюдь не означает, что само приращение процесса не зависит от времени (мы ни в коем случае не утверждаем, что для любых t, s, h справедливо $\eta_{t+h} - \eta_t = \eta_{s+h} - \eta_s$!), такое равенство означало бы движение частицы с постоянной скоростью. Как мы видели раньше, это не соответствует броуновскому движению. Аксиома 3) для броуновского движения, разумеется, отличается от соответствующей аксиомы для пуассоновского процесса. В данном случае движение частицы должно носить непрерывный характер, и приращение процесса за малое время должно быть малым. Тем более, это должно быть верно для третьей степени приращения. Поэтому будет использоваться следующая аксиома:

3) $\mathbb{E}|\eta_t|^3 = o(t)$, в частности, $\mathbb{E}|\eta_t|^3$ существует.

Отсюда следует существование и непрерывность среднего винеровского процесса η_t (по неравенству Гельдера

$$\mathbb{E}|\eta_{t+h} - \eta_t| = \mathbb{E}|\eta_h| \leq (\mathbb{E}|\eta_h|^3)^{1/3} = o(h^{1/3}),$$

точно так же существование и непрерывность второго момента, а следовательно, и дисперсии.

3. Среднее значение η_t .

По аксиоме 2) и свойству среднего

$$\mathbb{E}\eta_{t+s} = \mathbb{E}\eta_t + \mathbb{E}\eta_s,$$

то есть среднее – функция времени, удовлетворяющая условию аддитивности. Кроме того, эта функция непрерывна. Отсюда методами элементарного математического анализа выводится представление

$$\mathbb{E}\eta_t = mt,$$

где m – некоторая константа. (Опишем для полноты доказательство. Надо доказать, что для любого a

$$\mathbb{E}\eta_{at} = a\mathbb{E}\eta_t. \quad (1)$$

Используя 2), представления

$$\eta_t = \eta_{t/n} + (\eta_{2t/n} - \eta_{t/n}) + \dots + (\eta_t - \eta_{(n-1)t/n}),$$

$$\eta_{kt/n} = \eta_{t/n} + (\eta_{2t/n} - \eta_{t/n}) + \dots + (\eta_{kt/n} - \eta_{(k-1)t/n})$$

и аддитивность среднего, доказывается равенство (1) для $a = k/n$. Переходом к пределу мы получаем (1) в произвольном случае).

4. Дисперсия η_t

Точно так же, но с использованием аддитивности дисперсии для независимых приращений процесса (свойство 1)), доказывается

$$\mathbb{D}\eta_{at} = a\mathbb{D}\eta_t, \quad (2)$$

Поэтому

$$\mathbb{D}\eta_t = \sigma^2 t.$$

Итак, мы ввели два параметра винеровского процесса – *коэффициент сноса* t и *коэффициент диффузии* σ^2 .

5. Распределение случайной величины η_t .

Мы найдем характеристическую функцию

$$\varphi_t(u) = \mathbb{E}e^{iu\eta_t},$$

отождествим ее с известной нам характеристической функцией нормального распределения и таким образом докажем, что η_t имеет нормальное распределение.

Из свойств характеристической функции и аксиом 1), 2) следует:

$$\begin{aligned} \varphi_{t+\Delta t}(u) &= \mathbb{E}e^{iu\eta_{t+\Delta t}} = \mathbb{E}e^{iu(\eta_{t+\Delta t} - \eta_t)} e^{iu\eta_t} = \mathbb{E}e^{iu\eta_{\Delta t}} \mathbb{E}e^{iu\eta_t} = \\ &= \varphi_{\Delta t}(u)\varphi_t(u). \end{aligned}$$

Далее мы используем разложение характеристической функции в ряд Тейлора в окрестности нуля с представлением остаточного члена в форме Лагранжа:

$$\varphi_{\Delta t}(u) = \varphi_{\Delta t}(0) + \varphi'_{\Delta t}(0)u + \varphi''_{\Delta t}(0)\frac{u^2}{2!} + \varphi^{(3)}_{\Delta t}(\theta u)\frac{u^3}{3!}.$$

Запишем это равенство, используя средние и дисперсии:

$$\varphi_{\Delta t}(u) = 1 + i\mathbb{E}\eta_{\Delta t}u + i^2\mathbb{E}\eta_{\Delta t}^2\frac{u^2}{2!} + \mathbb{E}\left(e^{i\theta u\eta_{\Delta t}}i^3\eta_{\Delta t}^3\right)\frac{u^3}{3!}.$$

Подставив значения для среднего и дисперсии, и учтя условие 3), оценку

$$\left|\mathbb{E}e^{i\theta u\eta_{\Delta t}}i^3\eta_{\Delta t}^3\right| \leq \mathbb{E}\left|e^{i\theta u\eta_{\Delta t}}i^3\eta_{\Delta t}^3\right| = \mathbb{E}|\eta_{\Delta t}|^3 = o(\Delta t)$$

и формулу $\mathbb{E}\xi^2 = \mathbb{D}\xi + (\mathbb{E}\xi)^2$, имеем:

$$\varphi_{\Delta t}(u) = 1 + im\Delta tu - \frac{1}{2}(\sigma^2\Delta t + m^2(\Delta t)^2)u^2 + o(\Delta t).$$

В итоге

$$\begin{aligned} \varphi_{t+\Delta t}(u) &= \varphi_t(u) \left(1 + im\Delta tu - \frac{1}{2}(\sigma^2\Delta t + m^2(\Delta t)^2)u^2 + o(\Delta t) \right), \\ \frac{\varphi_{t+\Delta t}(u) - \varphi_t(u)}{\Delta t} &= \varphi_t(u) \left(imu - \frac{1}{2}\sigma^2u^2 + \frac{o(\Delta t)}{\Delta t} \right). \end{aligned}$$

Мы пришли к уравнению в частных производных

$$\frac{\partial \varphi_t(u)}{\partial t} = \varphi_t(u) \left(imu - \frac{1}{2}\sigma^2u^2 \right).$$

Решение этого уравнения (константа определяется при $t = 0$) имеет вид

$$\varphi_t(u) = e^{imut - \frac{1}{2}\sigma^2 u^2 t}.$$

Это — характеристическая функция нормального распределения $N(mt, \sigma\sqrt{t})$.

6. Ковариация процессов Пуассона и Винера.

Важнейшими характеристиками любого случайного процесса, η_t , являются среднее процесса $\mathbb{E}\eta_t$ и ковариация процесса

$$K(t, s) = \mathbb{E}(\eta_t - \mathbb{E}\eta_t)(\eta_s - \mathbb{E}\eta_s).$$

Давайте подсчитаем эти характеристики для процессов Пуассона и Винера.

Процесс Пуассона.

Среднее распределение Пуассона совпадает с его параметром и с его дисперсией. Напомним эти выкладки:

$$\begin{aligned} \mathbb{E}\xi &= \sum_k k \mathbb{P}\{k\} = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = \lambda, \\ \mathbb{E}\xi^2 &= \sum_k k^2 \mathbb{P}\{k\} = \sum_{k=1}^{\infty} \{k(k-1) + k\} \frac{\lambda^k}{k!} e^{-\lambda} = \lambda^2 + \lambda, \\ \mathbb{D}\xi &= \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2 = \lambda. \end{aligned}$$

Отсюда следует, что $\mathbb{E}\eta_t = \lambda t$, $\mathbb{D}\eta_t = \lambda t$.

Пусть $s < t$. Имеем (мы используем независимость приращений процесса):

$$\begin{aligned} K(t, s) &= \mathbb{E}(\eta_t - \mathbb{E}\eta_t)(\eta_s - \mathbb{E}\eta_s) = \mathbb{E}[(\eta_t - \mathbb{E}\eta_t) - (\eta_s - \mathbb{E}\eta_s)](\eta_s - \mathbb{E}\eta_s) + \\ &\quad \mathbb{E}(\eta_s - \mathbb{E}\eta_s)(\eta_s - \mathbb{E}\eta_s) = \\ &= \mathbb{E}[(\eta_t - \mathbb{E}\eta_t) - (\eta_s - \mathbb{E}\eta_s)]\mathbb{E}(\eta_s - \mathbb{E}\eta_s) + \mathbb{D}(\eta_s) = \mathbb{D}(\eta_s). \end{aligned}$$

Итак, в случае процесса Пуассона $K(t, s) = \lambda \min(t, s)$.

В случае процесса Винера точно так же $K(t, s) = \sigma^2 \min(t, s)$.

7. Оценка параметров процесса Винера

Оценка параметра m

$$\eta_t/t \rightarrow m \text{ в среднем квадратическом, т. е. } \mathbb{E}|\eta_t/t - m|^2 \rightarrow 0.$$

Для проверки сходимости в среднем квадратичном ξ_t к константе a достаточно:

- i) равенство или даже сходимость средних,
- ii) сходимость дисперсий к нулю.

Действительно,

$$\begin{aligned} \mathbb{E}|\xi_t - a|^2 &= \mathbb{E}|(\xi_t - \mathbb{E}(\xi_t)) + (\mathbb{E}(\xi_t) - a)|^2 \leq 2\mathbb{E}|\xi_t - \mathbb{E}(\xi_t)|^2 + 2\mathbb{E}|\mathbb{E}(\xi_t) - a|^2 = \\ &= 2\mathbb{D}(\xi_t) + 2\mathbb{E}|\mathbb{E}(\xi_t) - a|^2 \rightarrow 0. \end{aligned}$$

Теперь посмотрим наш случай: $\mathbb{E}(\eta_t/t) = m$, $\mathbb{D}(\eta_t/t) = \sigma^2 t/t^2 \rightarrow 0$.

Оценка параметра σ^2 винеровского процесса

Теорема. Допустим, что мы наблюдаем винеровский процесс до момента времени T . Рассмотрим последовательность разбиений интервала наблюдения на мелкие куски:

$$\Delta^{(n)} = 0 = t_1^{(n)} < t_2^{(n)} < \dots < t_{m(n)}^{(n)} = T, \quad \text{diam}(\Delta^{(n)}) = \max \left\{ t_{i+1}^{(n)} - t_i^{(n)} \right\} \rightarrow 0.$$

С каждым таким разбиением связывается оценка

$$\widehat{\sigma^2}^{(n)} = \sum_i \left| \eta_{t_{i+1}^{(n)}} - \eta_{t_i^{(n)}} \right|^2 / T,$$

которая сходится по вероятности к параметру σ^2 .

Доказательство. Напомним, что такие оценки мы называли состоятельными. Проверим, что

$$\widehat{\sigma^2}^{(n)} \xrightarrow{\mathbb{P}} \sigma^2.$$

i) Сходимость среднего к оцениваемому параметру:

$$\begin{aligned} \mathbb{E} \left(\widehat{\sigma^2}^{(n)} \right) / T &= \sum_i \mathbb{E} \left(\eta_{t_{i+1}^{(n)}} - \eta_{t_i^{(n)}} \right)^2 / T = \\ &= \sum_i \left\{ \mathbb{D} \left(\eta_{t_{i+1}^{(n)}} - \eta_{t_i^{(n)}} \right) + \left(\mathbb{E} \left(\eta_{t_{i+1}^{(n)}} - \eta_{t_i^{(n)}} \right) \right)^2 \right\} / T = \\ &= \sum_i \left(\sigma^2 \left(t_{i+1}^{(n)} - t_i^{(n)} \right) + m \left(t_{i+1}^{(n)} - t_i^{(n)} \right)^2 \right) / T = \\ &= \sigma^2 T / T + \sum_i m \left(t_{i+1}^{(n)} - t_i^{(n)} \right)^2 / T \rightarrow \sigma^2. \end{aligned}$$

Здесь мы использовали очевидную оценку

$$\begin{aligned} \sum_i \left(t_{i+1}^{(n)} - t_i^{(n)} \right)^2 &\leq \max_i \left(t_{i+1}^{(n)} - t_i^{(n)} \right) \sum_i \left(t_{i+1}^{(n)} - t_i^{(n)} \right) = \\ &= \text{diam}(\Delta^{(n)}) T \rightarrow 0. \end{aligned} \tag{3}$$

ii) Для проверки сходимости дисперсий к нулю нам будет удобно ввести вместо процесса η_t центрированный процесс $\xi_t = \eta_t - mt$. Нам нужно доказать, что

$$\mathbb{D} \left(\sum_i \left| \xi_{t_{i+1}^{(n)}} + mt_{i+1}^{(n)} - \xi_{t_i^{(n)}} - mt_i^{(n)} \right|^2 \right)^2 \rightarrow 0.$$

Мы воспользуемся следующим тривиальным неравенством $\mathbb{D}(\xi + \eta) \leq 2\mathbb{D}\xi + 2\mathbb{D}\eta$ — следствием неравенства $(x + y)^2 \leq 2x^2 + 2y^2$. Также мы используем формулы для второго и четвертого момента стандартной нормальной случайной величины ξ : $\mathbb{E}\xi^2 = 1$, $\mathbb{E}\xi^4 = 3$, в результате $\mathbb{D}\xi^2 = 3 - 1 = 2$, тогда $\mathbb{E}\xi_t^2 = t$, $\mathbb{E}\xi_t^4 = 3t^2$, в результате $\mathbb{D}(\xi_t^2) = 2t^2$.

Имеем:

$$\mathbb{D}(\xi_t + mt)^2 = \mathbb{D}(\xi_t^2 + 2mt\xi_t).$$

Теперь мы можем все подсчитать:

$$\begin{aligned} \mathbb{D} \left(\sum_i \left[(\xi_{t_{i+1}^{(n)}} - \xi_{t_i^{(n)}})^2 + 2m(t_{i+1}^{(n)} - t_i^{(n)})(\xi_{t_{i+1}^{(n)}} - \xi_{t_i^{(n)}}) \right] \right) &= \\ \sum_i \mathbb{D} \left((\xi_{t_{i+1}^{(n)}} - \xi_{t_i^{(n)}})^2 + 2m(t_{i+1}^{(n)} - t_i^{(n)})(\xi_{t_{i+1}^{(n)}} - \xi_{t_i^{(n)}}) \right) &\leq \\ \leq 2 \sum_i \mathbb{D} \left((\xi_{t_{i+1}^{(n)}} - \xi_{t_i^{(n)}})^2 \right) + 2 \sum_i \mathbb{D} \left(2m(t_{i+1}^{(n)} - t_i^{(n)})(\xi_{t_{i+1}^{(n)}} - \xi_{t_i^{(n)}}) \right) &\leq \end{aligned}$$

$$\leq 2 \sum_i 2 \cdot 2\sigma^2 (t_{i+1}^{(n)} - t_i^{(n)})^2 + 2 \sum_i 4m^2 (t_{i+1}^{(n)} - t_i^{(n)})^2 (t_{i+1}^{(n)} - t_i^{(n)}).$$

Наше утверждение следует из (3).

Резюмируем. Для хорошей оценки параметра m процесса Винера надо наблюдать процесс как можно дольше. Для хорошей оценки параметра σ^2 достаточно наблюдать процесс конечное время, но измерять его как можно чаще и точнее.

Заметим, что из усиленного закона больших чисел следует сходимость почти наверное $\eta_{nt}/nt \rightarrow m$, $n \rightarrow \infty$. Позднее мы докажем чрезвычайно любопытным способом сходимость $\eta_t(\omega)/t \rightarrow m$, где $t \rightarrow \infty$, для почти всех элементарных исходов ω . Впрочем, если быть точными, предварительно процесс η_t нужно будет заменить на *стохастически эквивалентный* процесс. Ниже мы определим это понятие, без введения которого невозможно изучать свойства траекторий случайного процесса.

8. Совместные распределения процессов.

Считается, что мы знаем процесс η_t , если знаем все совместные распределения $\mathbb{P}_{\eta_{t_1}, \eta_{t_2}, \dots, \eta_{t_n}}$ для всех наборов моментов времени $\{t_1, t_2, \dots, t_n\}$. Эта задача легко решается для описанных выше процессов – пуассоновского и винеровского.

Рассмотрим *пуассоновский процесс* η_t и $\{t_1, t_2, \dots, t_n\}$. Упорядочим эти моменты, без ограничения общности мы можем считать, что это уже сделано и $t_1 < t_2 < \dots < t_n$. Так как η_t принимают целочисленные неотрицательные значения, знание совместного распределения означает знание всех вероятностей вида

$$\mathbb{P}\{\eta_{t_1} = k_1, \eta_{t_2} = k_2, \dots, \eta_{t_n} = k_n\}. \quad (*)$$

Очевидно, что $k_1 \leq k_2 \dots \leq k_n$. Перепишем событие в вероятности (*) в другом виде:

$$\begin{aligned} &\mathbb{P}\{\eta_{t_1} = k_1, \eta_{t_2} = k_2, \dots, \eta_{t_n} = k_n\} = \\ &= \mathbb{P}\{\eta_{t_1} = k_1, \eta_{t_2} - \eta_{t_1} = k_2 - k_1, \dots, \eta_{t_n} - \eta_{t_{n-1}} = k_n - k_{n-1}\}. \end{aligned}$$

Далее мы учтем стационарность распределений приращений процесса и независимость приращений. Имеем:

$$\begin{aligned} &\mathbb{P}\{\eta_{t_1} = k_1, \eta_{t_2} - \eta_{t_1} = k_2 - k_1, \dots, \eta_{t_n} - \eta_{t_{n-1}} = k_n - k_{n-1}\} = \\ &= \mathbb{P}\{\eta_{t_1} = k_1\} \mathbb{P}\{\eta_{t_2} - \eta_{t_1} = k_2 - k_1\} \dots \mathbb{P}\{\eta_{t_n} - \eta_{t_{n-1}} = k_n - k_{n-1}\} = \\ &\mathbb{P}\{\eta_{t_1} = k_1\} \mathbb{P}\{\eta_{t_2-t_1} = k_2 - k_1\} \dots \mathbb{P}\{\eta_{t_n-t_{n-1}} = k_n - k_{n-1}\} = \frac{(\lambda t_1)^{k_1}}{k_1!} e^{-\lambda t_1} \\ &\frac{(\lambda(t_2-t_1))^{k_2-k_1}}{(k_2-k_1)!} e^{-\lambda(t_2-t_1)} \dots \frac{(\lambda(t_n-t_{n-1}))^{k_n-k_{n-1}}}{(k_n-k_{n-1})!} e^{-\lambda(t_n-t_{n-1})}. \end{aligned}$$

Аналогично получаются совместные распределения в случае, когда η_t — *винеровский процесс*, но теперь нам надо получить совместную плотность

$$p_{\eta_{t_1}, \eta_{t_2}, \dots, \eta_{t_n}}(x_1, x_2, \dots, x_n).$$

Для упрощения записей, мы упростим ситуацию и будем рассматривать лишь процессы с нулевым сносом ($m = 0$). Опять используя стационарность распределений приращений процесса и независимость приращений, имеем:

$$\begin{aligned} &p_{\eta_{t_1}, \eta_{t_2} - \eta_{t_1}, \dots, \eta_{t_n} - \eta_{t_{n-1}}}(x_1, x_2, \dots, x_n) = \\ &= \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n \frac{1}{\sqrt{t_1(t_2-t_1) \dots (t_n-t_{n-1})}} \times \end{aligned}$$

$$\times e^{-\frac{1}{2} \left(\frac{x_1^2}{\sigma^2 t_1} + \frac{x_2^2}{\sigma^2(t_2 - t_1)} + \cdots + \frac{x_n^2}{\sigma^2(t_n - t_{n-1})} \right)}.$$

Далее производим преобразование координат и находим совместную плотность

$$p_{\eta_{t_1}, \eta_{t_2}, \dots, \eta_{t_n}}(x_1, x_2, \dots, x_n).$$

Чтобы все это проделать осмысленно, мы вспомним определение совместной плотности:

$$\mathbb{P}\{(\eta_{t_1}, \eta_{t_2}, \dots, \eta_{t_n}) \in B\} = \int \dots \int_B p_{\eta_{t_1}, \eta_{t_2}, \dots, \eta_{t_n}}(x_1, x_2, \dots, x_n) dx_1 \dots dx_n.$$

Мы можем выразить событие $\{(\eta_{t_1}, \eta_{t_2}, \dots, \eta_{t_n}) \in B\}$ в терминах случайных величин $\eta_{t_1}, \eta_{t_2} - \eta_{t_1}, \dots, \eta_{t_n} - \eta_{t_{n-1}}$, точнее,

$$\{(\eta_{t_1}, \eta_{t_2}, \dots, \eta_{t_n}) \in B\} = \{\eta_{t_1}, \eta_{t_2} - \eta_{t_1}, \dots, \eta_{t_n} - \eta_{t_{n-1}} \in \tilde{B}\},$$

где B и \tilde{B} связаны следующим соотношением: $(x_1, x_1 + x_2, \dots, x_1 + x_2 + \dots + x_n) \in B$ тогда и только тогда, когда $(x_1, x_2, \dots, x_n) \in \tilde{B}$. Теперь вспомним определение функции плотности:

$$\begin{aligned} & \int \dots \int_B p_{\eta_{t_1}, \eta_{t_2}, \dots, \eta_{t_n}}(x_1, x_2, \dots, x_n) dx_1 \dots dx_n = \\ &= \mathbb{P}\{(\eta_{t_1}, \eta_{t_2}, \dots, \eta_{t_n}) \in B\} = \mathbb{P}\{\eta_{t_1}, \eta_{t_2} - \eta_{t_1}, \dots, \eta_{t_n} - \eta_{t_{n-1}} \in \tilde{B}\} = \\ &= \int \dots \int_{\tilde{B}} p_{\eta_{t_1}, \eta_{t_2} - \eta_{t_1}, \dots, \eta_{t_n} - \eta_{t_{n-1}}}(x_1, \dots, x_n) dx_1 \dots dx_n = \\ & \quad \int_{(x_1, x_1 + x_2, \dots, x_1 + \dots + x_n) \in B} p_{\eta_{t_1}, \eta_{t_2} - \eta_{t_1}, \dots, \eta_{t_n} - \eta_{t_{n-1}}}(x_1, x_2, \dots, x_n) dx_1 \dots dx_n = \\ &= \int_{(y_1, y_2, \dots, y_n) \in B} p_{\eta_{t_1}, \eta_{t_2} - \eta_{t_1}, \dots, \eta_{t_n} - \eta_{t_{n-1}}}(y_1, y_2 - y_1, \dots, y_n - y_{n-1}) dy_1 \dots dy_n. \end{aligned}$$

В предпоследнем интеграле мы заменили $y_k = x_1 + x_2 + \dots + x_k$.

Окончательно получаем

$$\begin{aligned} & p_{\eta_{t_1}, \eta_{t_2}, \dots, \eta_{t_n}}(x_1, x_2, \dots, x_n) = \\ &= \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n \frac{1}{\sqrt{t_1(t_2 - t_1) \dots (t_n - t_{n-1})}} \\ & \times e^{-\frac{1}{2} \left(\frac{x_1^2}{\sigma^2 t_1} + \frac{(x_2 - x_1)^2}{\sigma^2(t_2 - t_1)} + \dots + \frac{(x_n - x_{n-1})^2}{\sigma^2(t_n - t_{n-1})} \right)}. \end{aligned}$$

9. Согласованность совместных распределений.

Мы не заметили того, что отошли от первоначально предложенного определения случайного процесса. При определении как пуассоновского, так и винеровского процесса мы не использовали и не вводили никакого вероятностного пространства. Используя аксиомы, мы лишь вывели совместные распределения обоих процессов. Разумеется, если случайный процесс задавать на вероятностном пространстве, то для любого конечного набора моментов времени мы будем иметь совместное распределение соответствующих случайных величин ξ_t . Более того, эти распределения будут иметь свойство *согласованности*: вероятность $\mathbb{P}_{\xi_{t_1}, \dots, \xi_{t_n}}$ можно получить проектированием распределения $\mathbb{P}_{\xi_{t_1}, \dots, \xi_{t_n}, \xi_t}$, а именно:

$$\mathbb{P}_{\xi_{t_1}, \dots, \xi_{t_n}}(B) = \mathbb{P}_{\xi_{t_1}, \dots, \xi_{t_n}, \xi_t}(B \times \mathbf{R}).$$

Обратное нужно обосновывать. Но для этого сначала нужно проверить согласованность распределений, полученных нами выше. Мы ограничимся рассмотрением лишь стандартного винеровского процесса $w(t)$ (см. ниже) и лишь случаем трех моментов времени. Мы покажем, как осуществляется проверка согласованности лишь для совместных распределений $\mathbb{P}_{w(s), w(t), w(u)}$ и $\mathbb{P}_{w(s), w(u)}$, где $s < t < u$. Как мы уже знаем, для этого достаточно проверить тождество

$$\int_{-\infty}^{\infty} p_{w(s), w(t), w(u)}(x, y, z) dy = p_{w(s), w(u)}(x, z).$$

Наша задача упрощается тем, что мы имеем дело с функциями плотности, поэтому достаточно проверить равенство лишь для двух функций переменных x и z без учета постоянных множителей. Итак, зайдемся квадратичной формой под знаком экспоненты:

$$\begin{aligned} \frac{(x-y)^2}{t-s} + \frac{(z-y)^2}{u-t} &= \frac{x^2}{t-s} + \frac{z^2}{u-t} + \frac{(u-s)y^2}{(u-t)(t-s)} + 2\frac{xy}{t-s} + 2\frac{zy}{u-t} = \\ &= \frac{x^2}{t-s} + \frac{z^2}{u-t} + \frac{u-s}{(u-t)(t-s)} \left(y - \frac{u-t}{u-s}x - \frac{t-s}{u-s}z \right)^2 - \\ &\quad - \frac{u-t}{(u-s)(t-s)}x^2 - \frac{t-s}{(u-s)(u-t)}z^2 - 2\frac{1}{u-s}xz. \end{aligned}$$

После интегрирования по переменной $y - \frac{u-t}{u-s}x - \frac{t-s}{u-s}z$ под знаком экспоненты остается лишь квадратичная форма

$$\begin{aligned} \frac{x^2}{t-s} + \frac{z^2}{u-t} - \frac{u-t}{(u-s)(t-s)}x^2 - \frac{t-s}{(u-s)(u-t)}z^2 - 2\frac{1}{u-s}xz &= \\ &= x^2 \left(\frac{1}{t-s} - \frac{u-t}{(u-s)(t-s)} \right) + z^2 \left(\frac{1}{u-t} - \frac{t-s}{(u-s)(u-t)} \right) - 2\frac{1}{u-s}xz = \\ &= \frac{(z-x)^2}{u-s}, \end{aligned}$$

что и требовалось. (Рекомендуется выкладки с интегралом провести самим.)

10. Теорема Колмогорова.

Теорема. Пусть дано множество T и для каждого конечного набора $\{t_1, \dots, t_n\} \subset T$ задано распределение вероятностей $\mathbb{P}_{t_1, \dots, t_n}$ на борелевской σ -алгебре пространства \mathbf{R}^n (которое правилонее обозначать $\mathbf{R}^{\{t_1, \dots, t_n\}}$), причем набор распределений согласован (в смысле предыдущего пункта). Тогда на пространстве \mathbf{R}^T задана σ -алгебра, а на ней такая вероятностная мера \mathbb{P} , что совместное распределение случайных величин $(x_t) \rightarrow x_{t_i}$, $i \leq n$, совпадает с $\mathbb{P}_{t_1, \dots, t_n}$ для каждого набора $\{t_1, \dots, t_n\} \subset T$.

Нам понадобится одно свойство (регулярность) любой вероятностной меры \mathbb{P} на борелевской σ -алгебре пространства \mathbf{R}^n . Вероятностная мера \mathbb{P} называется *регулярной*, если для любого борелевского множества B и любого $\varepsilon > 0$ существуют замкнутое, F , и открытое, G , множества, такие, что

$$F \subset B \subset G, \quad \mathbb{P}(G \setminus B) < \varepsilon, \quad \mathbb{P}(B \setminus F) < \varepsilon. \quad (*)$$

Для доказательства мы введем класс всех регулярных множеств B , которые удовлетворяют $(*)$ для всех $\varepsilon > 0$ и докажем, что этот класс совпадает с σ -алгеброй борелевских множеств. Легко показывается, что каждое замкнутое множество F является регулярным.

Действительно (разберитесь !),

$$F = \bigcap_n \{x : d(x, F) < 1/n\}, \quad \mathbb{P}\{x : d(x, F) < 1/n\} \searrow \mathbb{P}(F)$$

и все множества $\{x : d(x, F) < 1/n\}$ открыты. Свойство регулярности сохраняется при переходе к дополнениям. Покажем, что это свойство сохраняется при переходе к счетному объединению: если все B_n регулярны, то $\bigcup_n B_n$ регулярно. Легко строится открытое G для $\bigcup_n B_n$: пусть $G_n \supset B_n$ и $\mathbb{P}(G_n \setminus B_n) < \varepsilon/2^n$, тогда в качестве G можно взять $\bigcup_n G_n$. Построение F сложнее (так как объединение замкнутых множеств не обязано быть замкнутым): сначала мы берем такое n , что

$$\mathbb{P}\left(\bigcup_n B_n \setminus \bigcup_{i=1}^n B_i\right) < \varepsilon/2,$$

потом для каждого $i \leq n$ берем такое замкнутое $F_i \subset B_i$, что

$$\mathbb{P}(B_i \setminus F_i) < \varepsilon/2^{i+1}.$$

Мы можем положить $F = \bigcup_{i=1}^n F_i$.

Заметим, что теорема о регулярности верна для любого метрического пространства. Более того, множества F мы можем считать компактными, пересекая их, если нужно с шарами $\{x : \|x\| \leq n\}$ и используя то, что $\mathbb{P}(\{x : \|x\| \leq n\}) \nearrow 1$.

Теперь перейдем к доказательству самой теоремы Колмогорова. Введем в \mathbf{R}^T алгебру цилиндрических множеств, для этого будет удобно представление

$$\mathbf{R}^T = \mathbf{R}^{\{t_1, \dots, t_n\}} \times \mathbf{R}^{T \setminus \{t_1, \dots, t_n\}}.$$

Каждое цилиндрическое множество представляется в виде

$$C(t_1, \dots, t_n; B) = B \times \mathbf{R}^{T \setminus \{t_1, \dots, t_n\}},$$

где B — борелевское множество в $\mathbf{R}^{\{t_1, \dots, t_n\}}$. Используя язык элементарной геометрии, мы будем называть B основанием цилиндра, а $\mathbf{R}^{T \setminus \{t_1, \dots, t_n\}}$ — образующими. Очевидно, что цилиндры образуют алгебру множеств, например, мы можем представить

$$C(t_1, \dots, t_n; B_1) \cup C(s_1, \dots, s_m; B_2) = C(t_1, \dots, t_n, s_1, \dots, s_m; B_3).$$

На этой алгебре функций множеств (пока это не вероятностная мера) \mathbb{P} задается соотношением

$$\mathbb{P}(C(t_1, \dots, t_n; B)) := \mathbb{P}_{t_1, \dots, t_n}(B).$$

Аддитивность этой функции множеств следует из согласованности системы конечномерных мер, так как два цилиндра с разными образующими представляются как цилиндры с одним образующим:

$$C(t_1, \dots, t_n; B_1) = C(t_1, \dots, t_n, s_1, \dots, s_m; B_1 \times \mathbf{R}^{\{s_1, \dots, s_m\}}),$$

$$C(s_1, \dots, s_m; B_2) = C(s_1, \dots, s_m, t_1, \dots, t_n; B_2 \times \mathbf{R}^{\{t_1, \dots, t_n\}}),$$

после этого используется аддитивность $\mathbb{P}_{t_1, \dots, t_n, s_1, \dots, s_m}$.

Итак, нам нужно доказать счетную аддитивность \mathbb{P} , которая, как мы знаем, эквивалентна непрерывности. Рассмотрим последовательность цилиндров $C_n \searrow \emptyset$, которую нам будет удобно задать в виде $C_n = C(t_1, \dots, t_n; B_n)$. Предположим противное:

$$\mathbb{P}(C(t_1, \dots, t_n; B_n)) \searrow \varepsilon > 0.$$

Далее в силу регулярности выберем внутри каждого основания B_n компактное основание K_n так, что

$$\mathbb{P}[C(t_1, \dots, t_n; B_n) \setminus C(t_1, \dots, t_n; K_n)] < \varepsilon/2^{n+1}.$$

Однако последовательность цилиндров с компактными основаниями K_n не обязательно является убывающей. Сделаем из нее убывающую последовательность:

$$C(t_1, \dots, t_n; \tilde{K}_n) := \bigcap_{i \leq n} C(t_1, \dots, t_i; K_i).$$

Имеем:

$$\mathbb{P}[C(t_1, \dots, t_n; B_n) \setminus C(t_1, \dots, t_n; \tilde{K}_n)] \leq \sum_{i \leq n} \mathbb{P}[C(t_1, \dots, t_n; B_i) \setminus C(t_1, \dots, t_n; K_i)] \leq \sum_{i \leq n} \varepsilon/2^{i+1} \leq \varepsilon/2.$$

Поэтому $\mathbb{P}[C(t_1, \dots, t_n; \tilde{K}_n)] > \varepsilon/2$, а это влечет, что все множества $C(t_1, \dots, t_n; \tilde{K}_n)$ непусты. Далее доказывается, что пересечение этих множеств непусто, что приводит нас к противоречию. Для доказательства в каждом компакте \tilde{K}_n выберем по элементу $(x_{t_1}^{(n)}, \dots, x_{t_n}^{(n)}) \in \mathbf{R}^{\{t_1, \dots, t_n\}}$. Превратим эту точку в элемент \mathbf{R}^T , считая функцию на остальных элементах T равной 0. По построению сужение элементов $x^{(m)}$, $m \geq n$, на $\{t_1, \dots, t_n\}$ принадлежит \tilde{K}_n . Ввиду компактности \tilde{K}_1 , из последовательности $(x_{t_1}^{(n)})$ выделим сходящуюся подпоследовательность $(x_{t_1}^{(n_k)})$, далее из подпоследовательности $(x_{t_1}^{(n_k)}, x_{t_2}^{(n_k)})$ выделим сходящуюся подпоследовательность $(x_{t_1}^{(n_{k_l})}, x_{t_2}^{(n_{k_l})})$. Повторяя эту процедуру счетное число раз для векторов все большей размерности, а потом используя метод выделения диагональной последовательности, мы получим подпоследовательность $(x^{(n_s)})$, сужение которой на каждое множество $\{t_1, \dots, t_n\}$ сходится. Обозначим поточечный предел нашей подпоследовательности через x . Тогда в силу компактности всех \tilde{K}_n , $x \in C(t_1, \dots, t_n; \tilde{K}_n)$ для всех n .

11. Парадоксы пуассоновского процесса. Случайные моменты остановки. Марковость.

В теории пуассоновского процесса большое значение имеют случайные величины τ_n — моменты n -й поломки станка. Из вероятностных соображений становится ясной независимость случайных величин $\tau_1, \tau_2 - \tau_1, \dots, \tau_n - \tau_{n-1}$. Но как строго обосновать этот интуитивно очевидный факт, ведь случайные величины $\tau_n - \tau_{n-1}$ вроде бы зависят от значений процесса в континуальное число моментов времени. Для упрощения обозначений мы ограничимся доказательством независимости двух первых случайных величин. Сначала вычислим совместную плотность p_{τ_1, τ_2} . Для упрощения записи будем считать $\lambda = 1$. Имеем:

$$\begin{aligned} \mathbb{P}(\{\tau_1 > x, \tau_2 > y\}) &= \mathbb{P}[\{\eta_x = 0\} \{\eta_y = 0\} + \{\eta_y = 1\}] = \\ &= \mathbb{P}[\{\eta_y = 0\}] + \mathbb{P}[\{\eta_x = 0\} (\{\eta_y - \eta_x = 1\})] = e^{-y} + e^{-x}(y-x)e^{-(y-x)} = \\ &= e^{-y}(1+y-x). \end{aligned}$$

Тогда

$$p_{\tau_1, \tau_2}(x, y) = \frac{\partial^2}{\partial x \partial y} e^{-y}(1+y-x)I_{\{0 < x < y\}} = e^{-y}I_{\{0 < x < y\}}$$

Далее с помощью замены переменных мы посчитаем совместную плотность случайных величин τ_1 и $\tau_2 - \tau_1$, она имеет вид

$$p_{\tau_1, \tau_2 - \tau_1}(x, y) = e^{-(x+y)}I_{\{x > 0\}}I_{\{y > 0\}} = p_{\tau_1}(x)p_{\tau_2 - \tau_1}(y).$$

Отсюда следует независимость и одинаковая распределенность случайных величин τ_1 и $\tau_2 - \tau_1$. Аналогично проверяется, что все случайные величины $\tau_k - \tau_{k-1}$ имеют одно и то же распределение и независимы.

Парадокс, связанный с пуассоновским процессом. Если мы зафиксируем большой момент времени t (так что с большой вероятностью он больше момента τ_1 и рассмотрим следующий момент поломки τ , то случайная величина $\tau - t$ в силу однородности по времени имеет то же распределение, что и любое $\tau_k - \tau_{k-1}$. Обозначим через $\tilde{\tau}$ момент поломки, предшествующий t

(или 0, если ранее поломок не было). Ясно, что случайная величина $t - \tilde{\tau}$ неотрицательна и не тождественный ноль. Поэтому распределение случайной величины $t - \tilde{\tau}$ сдвинуто в положительную сторону по сравнению с распределением $\tau_k - \tau_{k-1}$. В частности, у этого распределения больше среднее (при больших t оно почти вдвое больше). Между тем $t - \tilde{\tau}$ — это тоже время между двумя последовательными поломками, между моментами которых находится фиксированный момент времени t .

§31. Процессы массового обслуживания

1. Процессы массового обслуживания. Общая модель.

Обозначим снова через $p_n(t)$ вероятность наличия n элементов, описываемых нашим процессом. Классический однородный по времени процесс массового обслуживания задается бесконечной системой дифференциальных уравнений

$$p_n(t)' = -(\lambda + \mu)p_n(t) + \lambda p_{n-1}(t) + \mu p_{n+1}(t), n = 0, 1, 2, \dots, (*)$$

где характеристика λ называется *интенсивностью* заказов, а μ — *интенсивностью* исполнения заказов. В самой общей ситуации эти характеристики зависят от t и n . Метод исследования вероятностей $p_n(t)$ основан на использовании производящих функций. Эта задача лучше решается в следующем пункте, но лишь для частного случая. В общем случае системы уравнений (*) для производящей функции получаются уравнения в частных производных, содержащие производные высших порядков.

2. Преобразование Лапласа и производящая функция.

Преобразование Лапласа распределения \mathbb{P} неотрицательной случайной величины — функция неотрицательного параметра λ , задаваемая соотношением

$$\psi_{\mathbb{P}}(\lambda) = \int_{[0, \infty)} e^{-\lambda x} d\mathbb{P}(x).$$

Заметим, что для не обязательно неотрицательной величины интеграл может расходиться. Как и для характеристической функции, мы оперируем также с преобразованием Лапласа неотрицательной случайной величины — $\psi_{\xi}(\lambda) = \mathbb{E}e^{-\lambda\xi}$. Очевидны свойства преобразования Лапласа: i) $\psi(0) = 1$, ii) сама функция и все ее производные монотонны (такие функции называются *строго монотонными*), причем четные производные неотрицательны, нечетные производные неположительны, iii) преобразование Лапласа суммы независимых случайных величин является произведением преобразований Лапласа слагаемых. Замечательная теорема Бернштейна утверждает, что *любая функция на $[0, \infty)$ со свойствами i), ii) является преобразованием Лапласа некоторого распределения вероятностей на $[0, \infty)$ и однозначно определяет это распределение*.

Пример использования преобразования Лапласа. Как показать, что равномерное распределение на $[0, 1]$ не является сверткой двух одинаковых распределений? Надо вычислить преобразование Лапласа равномерного распределения ($\psi_{\xi}(\lambda = \frac{1}{\lambda}(1 - e^{-\lambda}))$), взять корень этого выражения и дифференцировать много раз с помощью пакета 'Математика', 12-я производная будет принимать не только положительные, но и отрицательные значения.

Для целочисленных неотрицательных случайных величин вместо преобразования Лапласа предпочтуют использовать *производящую функцию* $\pi_{\mathbb{P}}(z)$, заданную на $[0, 1)$ соотношением:

$$\pi_{\mathbb{P}}(z) = \sum_{n=0}^{\infty} \mathbb{P}\{n\} z^n.$$

Это преобразование получается из преобразования Лапласа подстановкой $z = e^{\lambda}$. Она также однозначно определяет распределение (здесь это очевидно) и удовлетворяет условию iii), однако $\pi_{\mathbb{P}}(0) = \mathbb{P}\{0\}$, $\lim_{z \searrow 1} \pi_{\mathbb{P}}(z) = 1$ (в этом можно легко убедиться).

В следующем пункте мы будем использовать производящую функцию геометрического распределения

$$\pi_g(z) = p + p(1-p)z + p(1-p)^2z^2 + \dots = \frac{p}{1 - (1-p)z}.$$

3. Процессы гибели и рождения.

Классический однородный по времени и по свойствам процесс гибели и рождения задается бесконечной системой дифференциальных уравнений

$$p_n(t)' = -n(\lambda + \mu)p_n(t) + (n-1)\lambda p_{n-1}(t) + (n+1)\mu p_{n+1}(t), n = 0, 1, 2, \dots,$$

где характеристика λ называется *интенсивностью* рождения, а μ — *интенсивностью* гибели. Предполагается, что вероятности перехода из состояния n в несоседние состояния мала по сравнению с вероятностями сохранения состояния или перехода в соседние состояния. Известно, что в ядерной реакции это не так, так как при распаде образуется от 1 до 3 нейтронов (состояние — число свободных нейтронов). Эту систему уравнений можно собрать с помощью понятия производящей функции —

$$\pi(z, t) = \sum_n p_n(t)z^n, 0 < z < 1.$$

Мы получаем следующее уравнение для производящей функции

$$\frac{\partial}{\partial t}\pi(z, t) = (\mu - (\lambda + \mu)z + \lambda z^2) \frac{\partial}{\partial z}\pi(z, t). \quad (*)$$

Решение уравнения ищется в виде произведения $\pi(z, t) = \varphi(t)\psi(z)$. Получаем дифференциальные уравнения для обеих функций:

$$\frac{\varphi(t)'}{\varphi(t)} = C, \frac{\psi(z)'}{\psi(z)} (\mu - (\lambda + \mu)z + \lambda z^2) = C.$$

Оба уравнения решаются, но общее решение получается в виде выпуклой комбинации этих решений, зависящих от C , возможно даже в виде интеграла по C по некоторой плотности от C . Впрочем, для получения значений вероятностей $p_n(t)$ производящую функцию нужно дифференцировать n раз по z . Это непросто, так как интерес (например, в задачах взрыва) представляют как раз вероятности больших n .

В данном простом случае общее решение получается даже без использования теории уравнений в частных производных. Действительно, уравнение

$$\frac{\partial}{\partial x}f(x, y) = \frac{\partial}{\partial u}f(x, y) \quad (**)$$

заменой переменных $u = x + y$, $v = x - y$. Представляя $f(x, y) = g(u, v)$ мы приходим к уравнению $\frac{\partial}{\partial v}g(u, v) = 0$, решением которого является произвольная дифференцируемая функция $g(u, v) = h(u)$, отсюда получаем $f(x, y) = h(x + y)$. Заметим для будущего, что, так как экспонента является взаимно однозначной функцией, мы можем представить $h(x + y) = \varphi(C_1 \cdot e^{C_2(x+y)})$. Уравнение $(*)$ сводится к $(**)$ заменой переменных: $\pi(t, z) = \tilde{\pi}(t, w)$, где $w = w(z)$ выбирается так, что $w'(z) = \frac{1}{(\lambda z - \mu)(z-1)}$, тогда $(\lambda z - \mu)(z-1)\frac{\partial}{\partial z}\pi(z, t) = \frac{\partial}{\partial w}\tilde{\pi}(z, t)$. Вычисляя неопределенный интеграл, с учетом сделанного выше замечания, получаем:

$$\pi(z, t) = \psi\left(\frac{\lambda z - \mu}{z-1}e^{-(\mu-\lambda)t}\right).$$

Функцию ψ мы вычисляем исходя из начального условия. А именно, если в момент $t = 0$ имеется n_0 особей, то $\pi(z, 0) = z^{n_0}$. Предварительно нужно убедиться в том, что

$$\frac{\mu - \frac{\lambda z - \mu}{z-1}}{\lambda - \frac{\lambda z - \mu}{z-1}} = z.$$

Поэтому

$$\psi(u) = \left(\frac{\mu - u}{\lambda - u} \right)^{n_0}.$$

В результате имеем:

$$\begin{aligned} \pi(z, t) &= \left(\frac{\mu - \frac{\lambda z - \mu}{z-1} e^{-(\mu-\lambda)t}}{\lambda - \frac{\lambda z - \mu}{z-1} e^{-(\mu-\lambda)t}} \right)^{n_0} = \left(\frac{(\mu z - \mu) - (\lambda z - \mu) e^{-(\mu-\lambda)t}}{(\lambda z - \lambda) - (\lambda z - \mu) e^{-(\mu-\lambda)t}} \right)^{n_0} = \\ &= \left(\frac{\mu(e^{-(\mu-\lambda)t} - 1) + z(\mu - \lambda e^{-(\mu-\lambda)t})}{(\mu e^{-(\mu-\lambda)t} - \lambda) - \lambda z(e^{-(\mu-\lambda)t} - 1)} \right)^{n_0}. \end{aligned}$$

Полное описание распределений мы можем получить при $n_0 = 1$, в этом случае наше распределение будет сверткой геометрического распределения с некоторым параметром $p_1(t)$ и распределения Бернулли с параметром $p_2(t)$. Для вычисления параметров представим производящую функцию как произведения производящей функции $\pi_g(z) = \frac{p_1(t)}{1 - (1 - p_1(t))z}$ геометрического распределения и производящей функции $\pi_b(z) = (1 - p_2(t)) + p_2(t)z$ распределения Бернулли. Вычисления проводятся в предположении $\mu \neq \lambda$. Сначала находим такое C , что

$$C\mu(e^{-(\mu-\lambda)t} - 1) + C(\mu - \lambda e^{-(\mu-\lambda)t}) = 1.$$

Имеем:

$$C = \frac{1}{\mu - \lambda} e^{(\mu-\lambda)t}, p_2(t) = \frac{\mu(e^{(\mu-\lambda)t} - \lambda)}{\mu - \lambda},$$

Деля $\pi(z, t)$ на $(1 - p_2(t)) + p_2(t)z$ и используя представление для $\pi_g(z)$, мы получаем

$$p_1(t) = \frac{(\mu - \lambda)e^{-(\mu-\lambda)t}}{\mu e^{-(\mu-\lambda)t} - \lambda}.$$

Теперь запишем вероятность n частиц в момент времени t :

$$\begin{aligned} p_n(t) &= p_2(t)p_1(t)(1 - p_1(t))^{n-1} + (1 - p_2(t))p_1(t)(1 - p_1(t))^n = \\ p_n(t) &= \frac{\mu(e^{(\mu-\lambda)t} - \lambda)}{\mu - \lambda} \cdot \frac{(\mu - \lambda)e^{-(\mu-\lambda)t}}{\mu e^{-(\mu-\lambda)t} - \lambda} \left(1 - \frac{(\mu - \lambda)e^{-(\mu-\lambda)t}}{\mu e^{-(\mu-\lambda)t} - \lambda} \right)^{n-1} + \\ &\quad + \left(1 - \frac{\mu(e^{(\mu-\lambda)t} - \lambda)}{\mu - \lambda} \right) \frac{(\mu - \lambda)e^{-(\mu-\lambda)t}}{\mu e^{-(\mu-\lambda)t} - \lambda} \left(1 - \frac{(\mu - \lambda)e^{-(\mu-\lambda)t}}{\mu e^{-(\mu-\lambda)t} - \lambda} \right)^n. \end{aligned}$$

§32. Свойства траекторий винеровского процесса

Для удобства мы в дальнейшем введем специальное обозначение для процесса броуновского движения с параметрами $m = 0$ и $\sigma = 1$. Такой процесс (в момент времени t) мы будем обозначать w_t , а иногда и $w(t)$.

1. Непрерывность траекторий винеровского процесса.

Этот факт является следствием более общей теоремы Колмогорова, дающей *достаточное* условие непрерывности почти всех траекторий случайного процесса.

Теорема. Пусть случайный процесс ξ_t удовлетворяет соотношению

$$\mathbb{E} |\xi_t - \xi_s|^\beta \leq C|t - s|^{1+\lambda} \quad \forall t, s, \text{ где } \lambda > 0, \beta > 0. \quad (1)$$

Тогда существует процесс $\tilde{\xi}_t$, стохастический эквивалентный ξ_t , с п.н. непрерывными траекториями.

Напомним, что *траекторией* процесса называется отображение

$$\tilde{\omega} : t \rightarrow \xi_t(\omega).$$

Свойство траекторий почти наверное означает, что вероятность множества всех ω , для которых это свойство выполняется, равна 1. При несчетном множестве значений времени t свойство траекторий может испортиться, если для каждого момента t процесс ξ_t изменится даже лишь на множестве вероятности 0. Разумеется, верно и обратное, в результате такого изменения траектории процесса могут стать хорошими.

Пример. Используем в качестве вероятностного пространства отрезок $[0, 1]$, причем события — борелевские подмножества $[0, 1]$, а вероятность — сужение на отрезок меры Лебега λ . Введем процесс

$$\xi_t = 0, \quad 0 \leq t \leq 1.$$

Разумеется, у такого процесса все траектории непрерывны. Рассмотрим эквивалентный процесс $\tilde{\xi}_t(\omega) = I_{\{t\}}(\omega)$. Для каждого $\omega \in [0, 1]$ траектория этого процесса везде нуль и непрерывна, за исключением точки $t = \omega$. Впрочем, в данном примере свойство непрерывности нарушается не очень сильно. Мы можем сделать траектории всюду разрывными, рассмотрев другой эквивалентный процесс

$$\tilde{\xi}_t(\omega) = I_{[0,1] \cap t}(\omega).$$

Поэтому для достижения нужного качества траекторий мы должны зарезервировать для себя право менять каждую случайную величину ξ_t на множестве вероятности 0, причем это множество зависит от t . Получающийся при этом формально новый процесс $\tilde{\xi}_t$ называется *стохастически эквивалентным* ξ_t . Важнейшее связывающее эти процессы свойство состоит в том, что эти процессы имеют одинаковые совместные распределения, т.е.

$$\mathbb{P}_{\xi_{t_1}, \xi_{t_2}, \dots, \xi_{t_n}} = \mathbb{P}_{\tilde{\xi}_{t_1}, \tilde{\xi}_{t_2}, \dots, \tilde{\xi}_{t_n}}.$$

Действительно, при переходе от процесса к эквивалентному процессу на множествах нулевой вероятности меняются лишь случайные величины $\xi_{t_1}, \xi_{t_2}, \dots, \xi_{t_n}$, вероятность объединения нулевых множеств также равна нулю. Для простоты мы будем считать, что $t \in [0, 1]$.

Доказательство теоремы Колмогорова. Идея конструкции состоит в следующем: мы рассматриваем процесс ξ_t лишь для $t = \frac{k}{2^n}$, где k, n — натуральные числа. Такие числа t называются двоично-рациональными, их множество мы обозначим через S . Далее мы считаем траектории

$$\tilde{\omega} : s \rightarrow \xi_s(\omega), s \in S.$$

заданными пока только на S . Потом доказываем, что для всех $\omega \in \Omega_1$, где $\mathbb{P}\Omega_1 = 1$, эти функции равномерно непрерывны на S . Далее продолжаем каждую такую $\tilde{\omega}$ на все $[0, 1]$, а значение этой функции в точке t обозначаем $\tilde{\xi}_t(\omega)$ (теперь это обозначение относится и к двоично рациональным t). Очевидно, что у нового процесса $\tilde{\xi}_t$ траектории непрерывны для всех $\omega \in \Omega_1$. Но нужно показать, что

$$\mathbb{P}\left\{\tilde{\xi}_t = \xi_t\right\} = 1 \quad \forall t, \tag{2}$$

одновременно мы покажем измеримость функций $\tilde{\xi}_t$. Выберем последовательность $s_n \in S$, сходящуюся к t . По неравенству Чебышева и (1)

$$\xi_{s_n} \xrightarrow{\mathbb{P}} \xi_t.$$

Согласно известному соотношению между сходимостью по вероятности и почти наверное, некоторая подпоследовательность ξ_{s_n} сходится почти наверное. Будем считать, что

$$\xi_{s_n} \xrightarrow{\text{П.Н.}} \xi_t.$$

Таким образом, $\xi_{s_n}(\omega) \rightarrow \xi_t(\omega)$ для всех ω из некоторого множества Ω_1 вероятности 1. В то же время по построению

$$\xi_{s_n}(\omega) \rightarrow \tilde{\xi}_t(\omega) \text{ для } \omega \in \Omega_2,$$

где множество Ω_2 имеет вероятность 1. Таким образом, $\tilde{\xi}_t(\omega) = \xi_t(\omega)$ для всех ω из пересечения $\Omega_1 \cap \Omega_2$ двух множеств с единичной вероятностью.

Теперь зайдемся процессом на S . Мы будем использовать лемму Бореля — Кантелли, согласно которой сходимость ряда вероятностей $\mathbb{P}A_n$ влечет принадлежность $\omega \in \Omega$ с вероятностью 1 лишь конечному числу A_n .

Применим к (2) неравенство Чебышева, имеем:

$$\mathbb{P}\left\{|\xi_{(k+1)/2^n} - \xi_{k/2^n}| \geq q^n\right\} \leq C \cdot 2^{-n-n\lambda} q^{-n\beta}.$$

Выберем $q = 2^{-\lambda/2\beta} < 1$, тогда

$$\mathbb{P}\left\{|\xi_{(k+1)/2^n} - \xi_{k/2^n}| \geq q^n\right\} \leq C \cdot 2^{-n} r^n, \quad (3)$$

где $r = 2^{-\lambda/2} < 1$.

Из (3) следует, что при любом n

$$\begin{aligned} \mathbb{P}\left\{\max_k |\xi_{(k+1)/2^n} - \xi_{k/2^n}| \geq q^n\right\} &= \mathbb{P}\left\{\bigcup_k \{|\xi_{(k+1)/2^n} - \xi_{k/2^n}| \geq q^n\}\right\} \leq \\ &\leq \sum_{k=0}^{2^n-1} \mathbb{P}\left\{|\xi_{(k+1)/2^n} - \xi_{k/2^n}| \geq q^n\right\} \leq 2^n \cdot C \cdot 2^{-n} r^n = C \cdot r^n. \end{aligned}$$

Ряд из этих вероятностей сходится, по лемме Бореля — Кантелли с вероятностью 1, начиная с некоторого $n = n(\omega)$ имеем

$$|\xi_{(k+1)/2^n}(\omega) - \xi_{k/2^n}(\omega)| < q^n. \quad (4)$$

Итак, будем считать, что ω удовлетворяет условию (4), и докажем равномерную непрерывность функции $t \rightarrow \xi_t(\omega)$ на множестве двоично рациональных t . Так как $q < 1$, то $q^n \rightarrow 0$, более того, ряд $\sum_n q^n$ сходится. Пусть теперь нам дано число $\varepsilon > 0$, мы должны выбрать для него $\delta > 0$ так, что $|s - t| \leq \delta$ влечет $|\xi_s(\omega) - \xi_t(\omega)| < \varepsilon$. Выберем сначала такое натуральное n , что $n > n(\omega)$ и

$$\sum_{i \geq n} q^i < \varepsilon/4.$$

Примем $\delta = 1/2^n$, и пусть $|s - t| \leq 1/2^n$. Если бы s и t были двоично рациональными точками вида $k/2^n$, то мы имели бы либо $s = t$, либо $s = m/2^n$, $t = (m+1)/2^n$ (если $t > s$), согласно (4) мы получили бы

$$|\xi_s(\omega) - \xi_t(\omega)| < q^n < \varepsilon/4.$$

Но возможна ситуация, когда s и t имеют вид $l/2^m$ и $r/2^p$, где $m > n$ и $p > n$. Тогда мы приближаем точку s соседним числом s_1 в разбиении на 2^m отрезков, которое имеет уже вид $u/2^{m-1}$, где $2u = l \pm 1$ (при этом $|\xi_s(\omega) - \xi_{s_1}(\omega)| < q^{m-1}$), далее s_1 приближаем числом s_2 вида $v/2^{m-2}$ и т. д., вплоть до числа \tilde{s} вида $k/2^n$. Очевидно, что $|s - \tilde{s}| < 1/2^n$ и ($s = s_0$)

$$|\xi_s(\omega) - \xi_{\tilde{s}}(\omega)| \leq \sum_{i=1}^x |\xi_{s_i}(\omega) - \xi_{s_{i-1}}(\omega)| \leq \sum_{i=n}^{\infty} q^i < \varepsilon/4.$$

Аналогичное приближение \tilde{t} мы находим для t : $|t - \tilde{t}| < 1/2^n$ и $|\xi_t(\omega) - \xi_{\tilde{t}}(\omega)| \leq \varepsilon/4$. Также по построению $|\tilde{s} - \tilde{t}| \leq 2/2^n$. В итоге

$$|\xi_s(\omega) - \xi_t(\omega)| \leq |\xi_s(\omega) - \xi_{\tilde{s}}(\omega)| + |\xi_{\tilde{s}}(\omega) - \xi_{\tilde{t}}(\omega)| + |\xi_{\tilde{t}}(\omega) - \xi_t(\omega)| \leq$$

$$\leq \varepsilon/4 + \varepsilon/4 + \varepsilon/2 = \varepsilon,$$

что и требовалось. Дальнейшие рассуждения в доказательстве теоремы Колмогорова приведены выше.

Следствие. Винеровский процесс эквивалентен процессу (этот процесс мы и будем обозначать w) с почти наверное непрерывными траекториями.

Проверим выполнение неравенства (1) для процесса w . Любопытно, что нас не устраивает $\beta = 2$, так как $\mathbb{E}|w(t) - w(s)|^2 = |t - s|$. Но $\beta = 4$ нас устраивает: $\mathbb{E}|w(t) - w(s)|^4 = 2|t - s|^2$, (1) выполняется.

2. Сходимость почти наверное для оценки параметра m

Теперь мы можем доказать сходимость почти наверное η_t/t к параметру m броуновского движения почти наверное при $t \rightarrow \infty$.

Для этого достаточно перейти к процессу $w(t)$ и доказать п.н. сходимость $w(t)/t$ к 0 ($t \rightarrow \infty$).

Оказывается, п.н. сходимость на ∞ следует из непрерывности траекторий в 0! Дело в том, что мы проведем в процессе $w(t)$ замену времени и рассмотрим вместо него процесс $tw(1/t)$. Очевидно, этот процесс также является центрированным. Проверим, что ковариация этого процесса совпадает с ковариацией винеровского процесса:

$$\mathbb{E}(tw(1/t)sw(1/s)) = ts \min\{1/t, 1/s\} = \min\{t, s\}.$$

Но мы знаем (см. многомерные характеристические функции), что совместное нормальное распределение однозначно определяется средними и вторыми моментами. Таким образом, совместные распределения нового процесса те же, что и у винеровского процесса $w(\cdot)$. Такое свойство винеровского процесса называется *авторегрессионностью*. Но тогда $tw(1/t)$ — винеровский процесс, и его траектории непрерывны, в частности, в нуле. Поэтому $tw(1/t)$ почти наверное сходится к нулю при $t \rightarrow 0$ (т. е. при $1/t \rightarrow \infty$).

Замечание 1. Свойство авторегрессионности разумеется не имеет общий процесс броуновского движения ξ_t с ненулевым m . Действительно,

$$\mathbb{E}(t\xi_{1/t}) = t \cdot \frac{m}{t} = m \neq mt = \mathbb{E}\xi_t.$$

Замечание 2. Заметим, что для п.н. сходимости случайного процесса нет эквивалентности двух определений предела — на языке $\varepsilon - \delta$ и на языке последовательностей. При доказательстве $\varepsilon - \delta$ определения мы имеем дело с несчетным набором множеств нулевой меры, мера объединения которых может быть не нуль.

Замечание 3. 'Улучшить' теорему Колмогорова, убрав в формулировке строго положительное λ , нельзя. Центрированный пуассоновский процесс η_t , как и винеровский процесс, удовлетворяет соотношению $\mathbb{E}|\eta(t) - \eta)|^2 = |t - s|$, но его траектории не могут быть непрерывными, они терпят разрыв в момент каждой поломки.

3. Случай сходимости почти наверное к параметру σ^2 .

4. Недифференцируемость траекторий винеровского процесса.

В доказательстве мы опять используем свойство авторегрессионности винеровского процесса: оказывается процесс $tw(\frac{1}{t})$ стохастически эквивалентен процессу $w(t)$.

Далее мы выясним, что означает дифференцируемость функции хотя бы в одной точке. Если $w(t)(\omega)$ дифференцируема в точке s , то

$$|w(t) - w(s)| < C(s)|t - s|$$

для всех $t \in (s, s+\varepsilon(s))$ в некоторой окрестности s . Разобьем отрезок, на котором задан процесс, на n частей, где $1/n$ меньше $\varepsilon/5$. Тогда для некоторого $l = l(s) > C(s)$, которое нам удобно считать натуральным (множество наших усилий должно быть счетным), для наименьшего $\frac{i}{n} \geq s$ мы имеем:

$$\left|w\left(\frac{i}{n}\right) - w(s)\right| \leq \frac{l}{n}, \left|w\left(\frac{i+1}{n}\right) - w(s)\right| \leq \frac{2l}{n}, \left|w\left(\frac{i+2}{n}\right) - w(s)\right| \leq \frac{3l}{n},$$

$$\left| w\left(\frac{i+3}{n}\right) - w(s) \right| \leq \frac{4l}{n}.$$

Отсюда следует, что

$$\left| w\left(\frac{j+1}{n}\right) - w\left(\frac{j}{n}\right) \right| \leq \frac{7l}{n},$$

для всех $j = i, i+1, i+2$. Отсюда мы выводим следующее описание множества всех траекторий, дифференцируемых хотя бы в одной точке отрезка $[0, 1]$:

$$M = \left\{ \omega : \bigcup_{l \in \mathbb{N}} \bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} \bigcup_{0 < i < n} \bigcap_{i \leq j \leq i+2} \left\{ \left| w\left(\frac{j+1}{n}\right) - w\left(\frac{j}{n}\right) \right| \leq \frac{7l}{n} \right\} \right\}$$

Нам нужно доказать, что

$\mathbb{P}(M) = 0$. Нетрудно видеть, что для этого достаточно показать

$$\mathbb{P}(M_{l,m}) = 0 \text{ для всех } l, m \in \mathbb{N},$$

где

$$M_{l,m} = \left\{ \omega : \bigcap_{n \geq m} \bigcup_{0 < i < n} \bigcap_{i \leq j \leq i+2} \left\{ \left| w\left(\frac{j+1}{n}\right) - w\left(\frac{j}{n}\right) \right| \leq \frac{7l}{n} \right\} \right\}.$$

Для этого достаточно показать

$$\mathbb{P} \left\{ \omega : \bigcup_{0 < i < n} \bigcap_{i \leq j \leq i+2} \left\{ \left| w\left(\frac{j+1}{n}\right) - w\left(\frac{j}{n}\right) \right| \leq \frac{7l}{n} \right\} \right\} \rightarrow 0 \text{ (} n \rightarrow \infty \text{).}$$

Имеем:

$$\begin{aligned} & \mathbb{P} \left\{ \omega : \bigcup_{0 < i < n} \bigcap_{i \leq j \leq i+2} \left\{ \left| w\left(\frac{j+1}{n}\right) - w\left(\frac{j}{n}\right) \right| \leq \frac{7l}{n} \right\} \right\} \leq \\ & \leq n \mathbb{P} \left\{ \omega : \bigcap_{i \leq j \leq i+2} \left\{ \left| w\left(\frac{j+1}{n}\right) - w\left(\frac{j}{n}\right) \right| \leq \frac{7l}{n} \right\} \right\} = \\ & = n \mathbb{P} \left\{ \omega : \left\{ \left| w\left(\frac{1}{n}\right) \right| \leq \frac{7l}{n} \right\} \right\}^3 = n \mathbb{P} \left\{ \omega : \left\{ |w(1)| \leq \frac{7l}{\sqrt{n}} \right\} \right\}^3 \leq \\ & \leq n \cdot \frac{\text{Const}}{n^{3/2}} \rightarrow 0, \end{aligned}$$

что и требовалось.

5. Функциональные предельные теоремы для процесса w .

Итак, процесс w , где $0 \leq t \leq 1$ можно рассматривать как случайный элемент $\omega \rightarrow w(\cdot)(\omega)$ со значениями в пространстве $C[0, 1]$. В таком случае распределение \mathbb{P}_w — это вероятностная мера на борелевской σ -алгебре в пространстве $C[0, 1]$. В пространстве всех вероятностных мер на этой σ -алгебре вводится понятие слабой сходимости точно так же, как и в пространствах вероятностных мер на борелевских σ -алгебрах конечномерных пространств. А именно, \mathbb{P}_n слабо сходится к \mathbb{P} тогда и только тогда, когда

$$\int_{C[0,1]} f(x) d\mathbb{P}_n(x) \rightarrow \int_{C[0,1]} f(x) d\mathbb{P}(x)$$

для всех непрерывных ограниченных функций f на $C[0, 1]$. В рамках этой теории мы кратко изложим результаты, которые в научной литературе называются принципом инвариантности Донскера — Прохорова о слабой сходимости к распределению винеровского процесса распределений случайных ломаных. В более общей теореме Прохорова рассматриваются серии ξ_{ni} , $i \leq k(n)$, независимых случайных величин из теоремы Линдеберга.

- 2) $\mathbb{E}\xi_{ni} = 0$ для всех n и k ;
- 3) $\sum_{i=1}^{k(n)} \mathbb{E}\xi_{ni}^2 = 1$ для каждого n ;

По каждой серии вводится процесс $S^{(n)}$ на $[0, 1]$, который называется случайной ломаной. Обозначим

$$t_k^{(n)} = \sum_{i=1}^k \mathbb{E}\xi_{ni}^2, k \leq k(n),$$

$$\eta_k^{(n)} = \sum_{i=1}^k \xi_{ni}.$$

Процесс $S^{(n)}$ задается следующими соотношениями:

$$S^{(n)}(t) = \begin{cases} \eta_k^{(n)}, & \text{если } t = t_k^{(n)}, k = 0, 1, \dots, k(n) \\ \text{линеен на каждом отрезке } [t_k^{(n)}, t_{k+1}^{(n)}]. \end{cases}$$

На всякий случай напомним, что функция f задается линейно на отрезке $[a, b]$, если для всех $\lambda \in (0, 1)$ имеет место равенство

$$f(\lambda a + (1 - \lambda)b) = \lambda f(a) + (1 - \lambda)f(b).$$

Как видите, мы построили действительно случайную ломаную, траектории которой безусловно непрерывны, то есть принадлежат пространству $C[0, 1]$. Ю.В. Прохоров доказал теорему: *распределение процесса $S^{(n)}$ на пространстве $C[0, 1]$ слабо сходится к распределению винеровского процесса*. Заметим, что процесс случайного блуждания на $\{0, 1, \dots, n\}$ также можно интерпретировать как случайную ломаную на отрезке $[0, 1]$, если процесс нормировать так, чтобы он в момент времени 0 имел дисперсию 1, траектории процесса нарисовать как ломаные, как мы это делали на лекции, точки k перевести в точки k/n ,

Из теоремы Прохорова и следующей из нее сходимости интегралов от непрерывных функций можно вывести также результаты о сходимости вероятностей. Например, вероятность пересечения уровня для процесса случайного блуждания сходится к соответствующей вероятности для винеровского процесса. В частности, так можно доказать следующее любопытное равенство

$$\mathbb{P}\{\max_{0 \leq t \leq 1} w(t) \geq x\} = 2\mathbb{P}\{w(1) \geq x\}. \quad (1)$$

6. Броуновский мост и его использование в математической статистике.

Большой интерес для математической статистики представляет броуновский мост, который называется также условным винеровским процессом. Этот процесс можно ввести не совсем строго как

$$\tilde{w}(t) := \mathbb{E}\{w(t) | w(1) = 0\}.$$

Таким образом броуновский мост начинается в 0 и кончается в 0. Совместные распределения броуновского моста нормальны и их можно определить следующим соотношением:

$$p_{\tilde{w}(t), \tilde{w}(s)}(x, y) = \frac{p_{w(t), w(s), w(1)}(x, y, 0)}{p_{w(1)}(0)}.$$

Упражнение 3. Вычислите корреляционную функцию процесса \tilde{w} . Докажите, что она совпадает с корреляционной функцией процесса $w(t) - tw(1)$.

Особый интерес представляет вопрос о распределении случайной величины $\max_{0 \leq t \leq 1} |\tilde{w}(t)|$. Это распределение является предельным для распределения статистики Колмогорова — Смирнова (при числе наблюдений, сходящемся к ∞).

7. Сильно марковское свойство винеровского процесса.

В этом пункте мы хотим доказать формулу (5.1) без использования функциональной предельной теоремы. Для этого мы хотим доказать аналог принципа отражения для случайного блуждания, однако в данной ситуации неприменимы рассуждения типа взамно однозначного соответствия между двумя множествами. Вместо функциональных предельных теорем мы применим теоремы о слабой сходимости многомерных распределений, которые были постулированы в 14.2 со ссылкой на одномерные аналоги, но не доказаны. Начнем со определений фильтрации случайного процесса и момента остановки случайного процесса, которые нам понадобятся и в других параграфах. Эти определения мы вводим для процессов с непрерывным временем $[0, \infty)$. Аналогично эти понятия вводятся для процессов с дискретным временем $\{0\} \cup \mathbb{N}$ и для процессов с прошлым, заданных на $(-\infty, \infty)$.

Определение 1. Пусть для каждого $t \geq 0$ задана σ -алгебра $\mathfrak{F}_t \subset \mathfrak{A}$, причем $\mathfrak{F}_t \subset \mathfrak{F}_s$ при $t \leq s$. Такой набор σ -алгебр называется *фильтрацией*.

Определение 2. Случайный процесс ξ_t называется *согласованным с фильтрацией* (\mathfrak{F}_t), если для любого t случайная величина ξ_t измерима относительно \mathfrak{F}_t .

Определение 3. Фильтрация (\mathfrak{F}_t) называется *естественной фильтрацией* случайного процесса ξ_t , если \mathfrak{F}_t для каждого t является наименьшей σ -алгеброй, относительно которой измеримы все случайные величины ξ_s ($s \leq t$).

Определение 4. Отображение $\tau : \Omega \rightarrow [0, \infty)$ называется *моментом остановки относительно фильтрации* (\mathfrak{F}_t), если $\{\tau \leq t\} \in \mathfrak{F}_t$ для любого $t \geq 0$.

Пример. Рассмотрим случайное блуждание с бесконечным временем. В качестве Ω возьмем пространство всех бесконечных последовательностей из знаков + и -. На Ω строится σ -алгебра множеств, порожденная всеми цилиндрами, состоящими из последовательностей (,, ..., , ω ...), в каждом цилиндре фиксированы первые n знаков, а остальные знаки произвольны. Мера каждого цилиндра с фиксированными первыми n -знаками принимается равной $1/2^n$. Мы можем заменить знаки числами 1, тогда $\Omega \subset \mathbf{R}^{\mathbb{N}}$, конечномерные меры согласованы и по теореме Колмогорова существует вероятность на σ -алгебре, порожденной алгеброй $\bigcup_n \mathfrak{F}_n$.

Предложение 1. Пусть $\tau : (\Omega, \mathfrak{A}, \mathbb{P}) \rightarrow [0, \infty)$ — момент остановки относительно естественной фильтрации (\mathfrak{F}_t) винеровского процесса w_t , $0 \leq t < \infty$. Тогда функция $w_\tau : \omega \rightarrow w_{\tau(\omega)}(\omega)$ является случайной величиной, т.е. измерима относительно σ -алгебры событий \mathfrak{A} .

Доказательство. Представим τ как предел дискретных случайных величин $\tau^{(n)}$ (см. обозначение в §7. п.1). Напомним, что

$$\tau^{(n)} = \sum_{k \in \mathbb{Z}} \frac{k}{n} I_{\left\{ \tau \in \left[\frac{k}{n}, \frac{k+1}{n} \right) \right\}}.$$

и $\tau^{(n)}$ равномерно сходится к τ . Проверим, что $w_{\tau^{(n)}}$ является случайной величиной:

$$\{w_{\tau^{(n)}} \in B\} = \sum_k \{w_{k/n} \in B\} \left\{ \tau^{(n)} = \frac{k}{n} \right\} \in \mathfrak{A},$$

так как все множества под знаком суммы являются событиями. Согласно теореме Винера из п. 1 мы можем считать, что функции $\omega : t \rightarrow w_t(\omega)$ являются непрерывными. Поэтому

$$w_{\tau^{(n)}}(\omega) \rightarrow w_\tau(\omega) \text{ для всех } \omega \text{ при } n \rightarrow \infty.$$

Замечание. В доказательстве использовалось лишь, что τ — неотрицательная случайная величина.

Предложение 2. *Моментом остановки для винеровского процесса w на $[0, \infty)$ для любого $y > 0$ является*

$$\tau_y(\omega) = \begin{cases} \inf\{t : w_t(\omega) \geq y\}, & \text{если такие } t > 0 \text{ существуют,} \\ \infty, & \text{если таких } t > 0 \text{ не существует.} \end{cases}$$

Доказательство. Очевидно, что ввиду непрерывности траекторий

$$\{\tau_y \leq t\} = \{\sup_{s \leq t} w_s \geq y\} = \{\sup\{w_s : s \in \mathbf{Q} \cap [0, t]\} \geq y\} \in \mathfrak{F}_t$$

для всех $t \geq 0$. Теперь покажем, что $\mathbb{P}\{\tau_y = \infty\} = 0$.

Предположим противное: $\mathbb{P}\{\tau_y = \infty\} = \varepsilon > 0$. Построим ряд из строго положительных чисел $\sum_n \lambda_n < \varepsilon/2$. По этому ряду построим по индукции последовательность положительных чисел $t_n \uparrow \infty$ и последовательность отрицательных чисел $x_n \downarrow -\infty$. Число x_n выбираются по t_{n-1} и t_n так, что

$$\mathbb{P}\{w_{t_n} - w_{t_{n-1}} \leq x_n\} < \lambda_n,$$

а t_{n+1} выбирается из условия $|x_1| + \dots + |x_n| + y < \sqrt{t_{n+1} - t_n}$. Обозначим $A = \bigcup_n \{w_{t_n} - w_{t_{n-1}} \leq x_n\}$, $\mathbb{P}(A) < \varepsilon/2$. Легко видеть, что

$$\{\sup_t w_t \leq y\} \subset \{\sup_n w_{t_n} \leq y\} \subset A \cup \bigcap_n \{w_{t_{n+1}} - w_{t_n} \leq |x_1| + \dots + |x_n| + y\}.$$

Поэтому

$$\begin{aligned} \mathbb{P}\{\tau_y = \infty\} &\leq \mathbb{P}(A) + \prod_n \mathbb{P}\{w_{t_{n+1}} - w_{t_n} \leq |x_1| + \dots + |x_n| + y\} \leq \mathbb{P}(A) + \prod_n \mathbb{P}\left\{\frac{w_{t_{n+1}} - w_{t_n}}{\sqrt{t_{n+1} - t_n}} < 1\right\} \\ &\leq \mathbb{P}(A) + \prod_n \mathbb{P}\{w_1 < 1\} < \varepsilon/2, \end{aligned}$$

что противоречит нашему предположению.

Далее мы докажем *принцип отражения*, позволяющий решать многие задачи в теории винеровского процесса. Мы опять будем использовать представление момента остановки τ как предела $\tau^{(n)}$ со значениями вида k/n .

Теорема 1. *Пусть τ — момент остановки винеровского процесса w_t , $0 \leq t < \infty$, тогда новый процесс*

$$w(\tau)_t(\omega) = \begin{cases} w_t(\omega), & \text{если } t \leq \tau(\omega), \\ 2w_{\tau(\omega)} - w_t(\omega), & \text{если } t > \tau(\omega). \end{cases}$$

также является винеровским.

Доказательство. Доказать, что $w_t^{(\tau)}$ является винеровским означает доказать, что совместные распределения случайных величин $w(\tau)_{t_1}, \dots, w(\tau)_{t_m}$ те же, что и у винеровского процесса. В силу непрерывности траекторий процесса w имеет место п.н. сходимость $w_t^{(\tau^{(n)})} \rightarrow w_t^{(\tau)}$ для любого t . Поэтому достаточно доказать теорему для любого дискретного момента остановки $\tau^{(n)}$. Добавим к уже упорядоченному набору $\{t_1, \dots, t_m\}$ все дроби $k/n \leq t_m$, получим набор $S = \{s_1, \dots, s_p\}$, который мы также будем считать упорядоченным по возрастанию. Заметим, что нам достаточно доказать совпадение совместных распределений наборов

$$(w_{s_1}, w_{s_2} - w_{s_1}, \dots, w_{s_p} - w_{s_{p-1}}),$$

$$(w_{s_1}^{(\tau^{(n)})}, w_{s_2}^{(\tau^{(n)})} - w_{s_1}^{(\tau^{(n)})}, \dots, w_{s_p}^{(\tau^{(n)})} - w_{s_{p-1}}^{(\tau^{(n)})}).$$

Обозначим $A_k = \{\tau^{(n)} = \frac{k}{n}\}$. Так как произведения одномерных множеств в пространстве \mathbf{R}^p порождают σ -алгебру (см. доказательство теоремы 6.4) нам достаточно проверить совпадение вероятностей событий.

Теорема 2.

Доказательство.

§33. Диффузионные процессы

1. Введение. Условные вероятности перехода.

Мы уже говорили о том, что марковские процессы $\eta(t)$ – это процессы, будущее которых при фиксированном настоящем не зависит от прошлого процесса. Таким образом, зная значение процесса $\eta(t)$, мы можем вычислить вероятности событий $\{\eta(t+h) \in B\}$, которые являются условными вероятностями $\mathbb{P}\{\eta(t+h) \in B | \eta(t) = x\}$. Марковскими процессами являются случайное блуждание, пуассоновский и винеровский процессы. Для пуассоновского процесса смысл условной вероятности прозрачен и мы имеем

$$\mathbb{P}\{\eta(t+h) \in B | \eta(t) = k\} = \frac{\mathbb{P}\{\eta(t+h) \in B, \eta(t) = k\}}{\mathbb{P}\{\eta(t) = k\}} = f(t, x, t+h, B).$$

Для винеровского процесса такая формула не подходит, так как

$$\mathbb{P}\{\eta(t) = x\} = 0 \text{ для любого } x.$$

Но у винеровского процесса можно ввести *условную плотность*. Сейчас мы введем это понятие. Пусть $p_{\xi,\eta}(x, y)$ — совместная плотность случайных величин ξ, η . Тогда у случайной величины ξ также имеется плотность, записываемая формулой:

$$p_\xi(x) = \int_{\mathbf{R}} p_{\xi,\eta}(x, y) dy.$$

Условная плотность задается формулой, аналогичной формуле для условной вероятности:

$$p_\eta(y | \xi = x) = \frac{p_{\xi,\eta}(x, y)}{p_\xi(x)}.$$

Итак, мы рассматриваем марковский случайный процесс $\eta(t)$ с непрерывным временем, где $0 \leq t \leq \infty$. Процесс может иметь счетное число состояний (то есть случайные величины $\eta(t)$ принимают лишь счетное число значений с ненулевой вероятностью, значения мы будем обозначать натуральными числами $i = 1, 2, 3, \dots$). В этом случае мы обозначим

$$P(s, i; t, j) = \mathbb{P}\{\eta(t) = j | \eta(s) = i\}.$$

Эти условные вероятности называются вероятностями перехода из состояния i в состояние j за время от s до t . В случае, когда случайные величины $\eta(t)$ имеют совместные плотности (а следовательно, множество значений никак не может быть счетным) мы обозначим плотность вероятностей перехода из состояния x в состояние y за время от s до t :

$$p(s, x; t, y) = p_{\eta(t)}(y | \eta(s) = x).$$

2. Уравнение Колмогорова — Чепмена.

Пусть $s < u < t$. Тогда

$$p(s, x; t, y) = \int_{-\infty}^{\infty} p(s, x; u, z) p(u, z; t, y) dz.$$

В непрерывном случае для доказательства уравнения Колмогорова — Чепмена нужно приложить определенные усилия. В дискретном случае счетного числа состояний все очень просто и мы доказали дискретный аналог этого равенства (пункт 28.1):

$$P(s, i; t, j) = \sum_k P(s, i; u, k)P(u, k; t, j). \quad (***)$$

Отметим также тождество, которое будет позднее использовано несколько раз:

$$\int_{-\infty}^{\infty} p(s, x; u, z)dz = 1. \quad (***)$$

3. Аксиомы диффузионных процессов.

В дальнейшем предполагается, что существуют такие функции $a(t, x)$ и $b(t, x)$, зависящие от времени и места, что для всех $\varepsilon > 0$ имеют место соотношения:

$$\int_{|z-x|>\varepsilon} p(t, x; t + \Delta t, z)dz = o(\Delta t), \quad (1)$$

$$\int_{|z-x|\leq\varepsilon} (z - x)p(t, x; t + \Delta t, z)dz = a(t, x)\Delta t + o(\Delta t), \quad (2)$$

$$\int_{|z-x|\leq\varepsilon} (z - x)^2 p(t, x; t + \Delta t, z)dz = b(t, x)\Delta t + o(\Delta t). \quad (3)$$

Легко проверяется, что процесс броуновского движения также удовлетворяет этим аксиомам с

$$a(t, x) = m, b(t, x) = \sigma^2.$$

Таким образом, диффузионный процесс можно рассматривать как броуновское движение в неоднородной среде, свойства которой со временем меняются.

4. Дифференциальные уравнения Колмогорова

Прямое уравнение Колмогорова

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial y}[a(t, y)p(s, x; t, y)] + \frac{1}{2}\frac{\partial^2}{\partial y^2}[b(t, y)p(s, x; t, y)].$$

Приведем здесь, кое-где не очень строгое, доказательство (из книги Ю.А. Розанова). Рассмотрим пробную функцию φ (т. е. бесконечно дифференцируемую функцию с компактным носителем)

$$\varphi(z) = \varphi(y) + \varphi'(y)(z - y) + \frac{1}{2}\varphi''(y)(z - y)^2 + o(z - y)^2. \quad (4)$$

Напомним, что равенство нулю интеграла от произведения всех пробных функций с данной функцией f влечет равенство почти всюду $f = 0$. Составим произведение (используя уравнение Чепмена — Колмогорова):

$$p(s, x, t + \Delta t, y) = \int_{-\infty}^{\infty} p(s, x; t, z)p(t, z, t + \Delta t, y)dz. \quad (*)$$

Теперь рассмотрим интеграл от производной по времени переходной плотности, умноженной на пробную функцию:

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (\varphi(y))[p(s, x; t + \Delta t, y) - p(s, x; t, y)]dy =$$

(учтем предыдущую формулу)

$$= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(y) p(s, x; t, z) p(t, z; t + \Delta t, y) dy dz - \int_{-\infty}^{\infty} \varphi(y) p(s, x; t, y) dy dz =$$

(далее мы поменяем ролями в первом интеграле переменные y и z и учтем, что $\int_{-\infty}^{\infty} p(t, y; t + \Delta t, z) dz = 1$)

$$= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\varphi(z) - \varphi(y)] p(s, x, t, y) p(t, y; t + \Delta t, z) dy dz =$$

(учитывая (1) и ограниченность φ , мы по z можем интегрировать лишь по некоторому интервалу $|z - y| < \varepsilon$, а ε выбираем сначала из условия малости $o(z - y)^2$ в (4))

$$\begin{aligned} &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{y-\varepsilon}^{y+\varepsilon} \left(\varphi'(y)(z - y) + \frac{1}{2} \varphi''(y)(z - y)^2 + o(z - y)^2 \right) p(s, x, t, y) p(t, y; t + \Delta t, z) dz dy = \\ &\quad = \int_{-\infty}^{\infty} \left[\varphi'(y) a(t, y) + \frac{1}{2} \varphi''(y) b(t, y) \right] p(s, x; t, y) dy. \end{aligned}$$

Далее каждое из слагаемых мы интегрируем по частям, первое слагаемое один раз и второе — два раза. Мы используем равенство нулю функции φ и ее производных вне некоторого интервала, поэтому слагаемые без интегралов зануляются. Напомним, что каждое интегрирование по частям меняет знак перед интегралом. Итак,

$$\int_{-\infty}^{\infty} \varphi(y) \frac{\partial}{\partial t} p(s, x; t, y) dy = \int_{-\infty}^{\infty} \varphi(y) \left\{ -\frac{\partial}{\partial y} [a(t, y) p(s, x; t, y)] + \frac{1}{2} \frac{\partial^2}{\partial y^2} [b(t, y) p(s, x; t, y)] \right\} dy.$$

Ввиду произвольности пробной функции φ , получаем уравнение.

Обратное уравнение Колмогорова

$$-\frac{\partial p(s, x; t, y)}{\partial s} = a(s, x) \frac{\partial p(s, x; t, y)}{\partial x} + \frac{1}{2} b(s, x) \frac{\partial^2 p(s, x; t, y)}{\partial x^2}.$$

Строгое доказательство вы можете найти в книге Ю.А. Розанова. Там используется аппарат пробных функций (из теории обобщенных функций). Благодаря использованию пробных функций, как и выше, можно обосновать использование аксиом (1-3). Здесь я попробую объяснить причины этого уравнения.

$$p(s, x; t, y) = \int_{-\infty}^{\infty} p(s, x; s + \Delta s, z) p(s + \Delta s, z; t, y) dz. \quad (5)$$

Разложим $p(s + \Delta s, z; t, y)$ в ряд Тейлора в окрестности точки $(s, z = x)$ по степеням Δs и $z - x$. Получаем:

$$\begin{aligned} &p(s + \Delta s, z; t, y) = \\ &= p(s, x; t, y) + \frac{\partial p}{\partial s}(s, x; t, y) \Delta s + \frac{\partial p}{\partial z}(s, x; t, y) (z - x) + \frac{1}{2} \frac{\partial^2 p}{\partial z^2}(s, x; t, y) (z - x)^2. \end{aligned}$$

Подставляем в (5), имеем (но все это очень нестрого!):

$$\begin{aligned} p(s, x; t, y) &= \int_{-\infty}^{\infty} p(s, x; s + \Delta s, z) p(s, x; t, y) dz + \\ &\quad + \int_{-\infty}^{\infty} p(s, x, s + \Delta s, z) \frac{\partial p}{\partial s}(s, x; t, y) \Delta s dz + \\ &\quad + \int_{-\infty}^{\infty} p(s, x; s + \Delta s, z) \frac{\partial p}{\partial x}(s, x; t, y) (z - x) dz + \int_{-\infty}^{\infty} p(s, x; s + \Delta s, z) \frac{1}{2} \frac{\partial^2 p}{\partial x^2}(s, x; t, y) (z - x)^2 dz. \end{aligned}$$

Используя аксиомы диффузионных процессов (это еще менее строго!), имеем (четырежды используем (***)�

$$\begin{aligned} p(s, x; t, y) &= p(s, x; t, y) + \\ &+ \frac{\partial p}{\partial s}(s, x; t, y)\Delta s + \frac{\partial p}{\partial x}(s, x; t, y)a(s, x)\Delta s + \frac{\partial^2 p}{\partial x^2}(s, x; t, y)b(s, x)\Delta s + o(\Delta s). \end{aligned}$$

Дальше сокращаем $p(s, x; t, y)$ в левой и правой части и все делим на Δs , получаем обратное уравнение Колмогорова.

Упражнение. Проверьте, что условная плотность винеровского процесса удовлетворяет аксиомам диффузионных процессов и обоим уравнениям Колмогорова.

§34. Стохастические дифференциальные уравнения

Винеровский процесс рассматривается также в радиотехнике и интерпретируется там как результат интегрирования помех — белого шума. В теории управления движущими аппаратами интегральное влияние случайной среды (например, порывы ветра, изменения давления) также считается винеровским процессом. Но при этом возникают и другие эффекты — в результате влияния среды меняется положение управляющих аппаратом систем (двигателя), и случайно накопившаяся ошибка может давать при этом дополнительную ошибку. Для описания такого рода явлений используется аппарат стохастических дифференциальных уравнений. Они вводятся с помощью различных стохастических интегралов.

Еще раз напомним, что для упрощения выкладок введем стандартный винеровский процесс $w(t)$ — это винеровский процесс с нулевым сносом $m = 0$ и единичным коэффициентом диффузии $\sigma = 1$. Произвольный винеровский процесс получается из него линейным преобразованием. Мы также изменим несколько обозначения и будем писать все процессы как функции времени — $\xi(t)$

1. Стохастические интегралы.

Стохастические интегралы, в которые входят случайные процессы $\xi(t)$, бывают трех видов:

$$\int_0^T \{g(t) + f(t)\xi(t)\} dt,$$

$$\int_0^T f(t)d\xi(t),$$

где $f(t)$ — это обычная функция, и самые сложные и случайные —

$$\int_0^T \eta(t)d\xi(t),$$

где ξ и η — два случайных процесса.

Стохастические интегралы определяются так же, как и обычные, то есть как предел интегральных сумм. Но при определении самых сложных интегралов возникают любопытные проблемы. Пределы интегрирования могут быть другими, но при сделанной выше записи видно, что в результате интегрирования возникают новые случайные процессы, $\zeta(T)$.

Интегральные суммы для $\int_0^T f(t)w(t)dt$ записываются очевидным образом:

$$\sum f(t_i)w(t_i)(t_{i+1} - t_i),$$

где $0 = t_0 < t_1 < \dots < t_{n-1} < t_n = T$.

Сходимость интегральных сумм легко доказывается для непрерывной функции f и даже в значительно более общих ситуациях. Проще всего доказывать сходимость в среднем квадратическом (то есть $\xi_n \rightarrow \xi$, если $|\mathbb{E}|\xi_n - \xi|^2 \rightarrow 0$). Сами интегральные суммы являются интегралами от кусочно постоянных случайных процессов. Знак \lim ниже означает предел интегральных сумм.

В данной ситуации случайная величина $\int_0^T f(t)w(t)dt$ также имеет нормальное распределение. Попробуем определить параметры этого распределения, то есть подсчитать среднее и дисперсию. Естественно, что все это мы будем делать не для интеграла общего вида, а лишь для интегральных сумм.

Итак,

$$\begin{aligned}\mathbb{E} \int_0^T (g(t) + f(t))w(t)dt &= \lim \mathbb{E} \left(\sum_i (g(t_i) + f(t_i))w(t_i)(t_{i+1} - t_i) \right) = \\ &= \lim \sum_i g(t_i)(t_{i+1} - t_i) = \int_0^T g(t)dt.\end{aligned}$$

Значительно хитрее считаются дисперсия или центральный смешанный момент. При ее подсчете мы вычитаем среднее, то есть убираем неслучайное слагаемое $g(t)$ из под знака интеграла.

$$\begin{aligned}\mathbb{E} \left\{ \int_0^T f(t)w(t)dt \int_0^T h(t)w(t)dt \right\} &= \\ &= \lim \mathbb{E} \left\{ \sum_i f(t_i)w(t_i)(t_{i+1} - t_i) \sum_j h(s_j)w(s_j)(s_{j+1} - s_j) \right\} = \\ &= \lim \sum_i \sum_j \mathbb{E} (f(t_i)w(t_i)(t_{i+1} - t_i)h(s_j)w(s_j)(s_{j+1} - s_j)) = \\ &= \lim \sum_i \sum_j f(t_i)h(s_j)K(t_i, s_j)(t_{i+1} - t_i)(s_{j+1} - s_j) = \\ &= \int_0^T \int_0^T f(t)h(s)K(t, s)dtds.\end{aligned}$$

Напомним, что в нашем случае $K(t, s) = \min(t, s)$. В частности,

$$\mathbb{D} \left(\int_0^T f(t)w(t)dt \right) = \int_0^T \int_0^T f(t)f(s) \min(t, s)dtds.$$

Если в приближении интегральными суммами использовать в обоих интегралах одно и то же разбиение, то проще вычисляется

$$\begin{aligned}\mathbb{D} \left(\int_0^T f(t)dw(t) \right) &= \\ &= \mathbb{E} \left(\int_0^T f(t)dw(t) \int_0^T f(t)dw(t) \right) =\end{aligned}$$

$$\begin{aligned}
&= \lim \sum_i \sum_j \mathbb{E} (f(t_i)(w(t_{i+1}) - w(t_i))f(t_j)(w(t_{j+1}) - w(t_j))) = \\
&= \lim \sum_i \sum_j \mathbb{E} (f(t_i)(w(t_{i+1}) - w(t_i))f(t_j)(w(t_{j+1}) - w(t_j))) = \\
&= \lim \sum_i f(t_i)^2 (t_{i+1} - t_i) = \int_0^T f^2(t) dt.
\end{aligned}$$

Проверьте выкладку. Мы использовали равенство нулю среднего процесса w и независимость приращений w .

Замечание. Кажется, что при определении интеграла $\int_0^T f(t)dw(t)$ мы можем использовать то обстоятельство, что процесс $w(t)$ можно рассматривать на каждом элементарном исходе ω отдельно как функцию $w(t)(\omega)$, а интеграл также записать для каждого ω отдельно как $\int_0^T f(t)w'(t)(\omega)dt$. Увы, это невозможно. Винер показал, что можно считать функции $t \rightarrow w(t)(\omega)$ непрерывными для всех ω , в то же время все эти функции являются недифференцируемыми ни при одном t (с вероятностью 1). Однако, используя такое разложение процесса по элементарным исходам, мы вполне можем вводить интегралы вида $\int_0^T f(t)w(t)dt$.

2. Интеграл Ито.

Наконец, самый сложный и нетривиальный случай стохастического интеграла — $\int_0^T \eta(t)dw(t)$, где $\eta(t)$ — случайный процесс. Возможны разные варианты введения этого стохастического интеграла. Вообще говоря, неясно, какой из этих способов наилучшим образом соответствует практическим задачам. Наиболее употребительным является *интеграл Ито*.

В определении интеграла Ито

$$\int_0^T \eta(t)dw(t) \tag{3}$$

процесс $\eta(t)$ должен иметь важное свойство *предсказуемости*. А именно, значения процесса $\eta(t)$ должны определяться значениями процесса $w(s)$ во все промежутки времени до t включительно, но не должны зависеть от приращений процесса $w(s)$ в будущем, при $s > t$. Таким образом, процесс $\eta(t)$ появляется в результате действия помех $w(t)$ в прошлом, а не в будущем.

Пример предсказуемого процесса — $\eta(t) = \int_0^t w(s)ds$.

Ито предложил определить интеграл

$$\int_0^T \eta(t)dw(t)$$

как предел интегральных сумм вида

$$\sum_{i=0}^n \eta(t_i)(w(t_{i+1}) - w(t_i)),$$

где, как обычно, $0 = t_0 < t_1 < \dots < t_{n-1} < t_n = T$ — разбиение Δ отрезка $[0, T]$. Но что произойдет, если мы рассмотрим другие интегральные суммы, например,

$$\sum_{i=0}^n \eta(t_{i+1})(w(t_{i+1}) - w(t_i)).$$

Изменение незначительно, и в теории интеграла Римана показывается, что изменение аргумента подинтегральной функции в интервале разбиения не меняет предел. Оказывается, что в нашем случае случайной подинтегральной функции и случайного дифференциала предел изменится. Рассмотрим, например, интеграл

$$\int_0^T w(t) dw(t).$$

Вычтем из интегральной суммы второго вида

$$\sum_{i=0}^n w(t_{i+1})(w(t_{i+1}) - w(t_i))$$

интегральную сумму Ито

$$\sum_{i=0}^n w(t_i)(w(t_{i+1}) - w(t_i)).$$

Разность равна

$$\sum_{i=0}^n (w(t_{i+1}) - w(t_i))^2 \quad (**)$$

и сходится, как мы уже доказали в пункте 29.4 (оценка параметров) не к 0, а к T .

3. Задача. Вычислить интеграл Ито

$$\int_0^T w(t) dw(t).$$

Задача решается искусственным приемом. По видимому, соображением, позволяющим догадаться до этого приема, служит то обстоятельство, что в неслучайной ситуации интеграл должен был бы равняться

$$(w^2(T) - w^2(0)) / 2.$$

Рассмотрим удвоенное значение этой величины и попробуем связать его с разбиением Δ :

$$\begin{aligned} w^2(T) - w^2(0) &= w^2(t_n) - w^2(t_0) = \\ &= [w^2(t_n) - w^2(t_{n-1})] + [w^2(t_{n-1}) - w^2(t_{n-2})] + \cdots + [w^2(t_2) - w^2(t_1)] + [w^2(t_1) - w^2(t_0)]. \end{aligned}$$

Каждое слагаемое $w^2(t_{i+1}) - w^2(t_i)$ в этой сумме можно представить в виде

$$2w(t_i)[w(t_{i+1}) - w(t_i)] + (w(t_{i+1}) - w(t_i))^2.$$

Первое слагаемое в этом представлении входит в интегральную сумму Ито. Второе слагаемое входит в сумму (**), предел которой известен. Таким образом, мы получаем представление

$$w^2(T) = \sum_{i=0}^{n-1} 2w(t_i)[w(t_{i+1}) - w(t_i)] + \sum_{i=0}^{n-1} (w(t_{i+1}) - w(t_i))^2.$$

Устремляя диаметр разбиения к 0 и переходя к пределу, получаем

$$w^2(T) = 2 \int_0^T w(t) dw(t) + T.$$

$$\int_0^T w(t) dw(t) = (w^2(T) - T)/2.$$

4. Существование интеграла Ито.

Теперь попытаемся объяснить – что такое интеграл Ито

$$\int_0^T \eta(t) dw(t)$$

и почему он существует для хороших подынтегральных процессов $\eta(t)$. Напомним, что процесс $\eta(t)$ предполагается зависящим от прошлого и настоящего процесса w . Кроме того, входящие в него случайные величины должны иметь дисперсию. Мы будем также считать процесс $\eta(t)$ непрерывным в среднем квадратическом. Интеграл Ито (как и любой другой интеграл) мы определяем как предел интегральных сумм

$$\sum_{i=0}^n \eta(t_i)(w(t_{i+1}) - w(t_i)),$$

где $0 = t_0 < t_1 < \dots < t_{n-1} < t_n = T$ – разбиение Δ отрезка $[0, T]$. Эта интегральная сумма на деле является интегралом кусочно-постоянного процесса

$$\eta_\Delta(t) = \sum_{i=0}^{n-1} I_{[t_i, t_{i+1})}(t)\eta(t_i).$$

Сходимость интегралов мы будем понимать в среднем квадратическом. Поэтому имеет смысл считать среднее и дисперсию интегральных сумм. Так как приращения процесса w не зависят от прошлого процесса w , они не зависят и от прошлого процесса $\eta(t)$. Поэтому

$$\mathbb{E} \left\{ \sum_{i=0}^n \eta(t_i)(w(t_{i+1}) - w(t_i)) \right\} = \sum_{i=0}^n \mathbb{E}\eta(t_i)\mathbb{E}(w(t_{i+1}) - w(t_i)) = \sum_{i=0}^n \mathbb{E}\eta(t_i) \times 0 = 0.$$

Дисперсия считается сложнее.

$$\begin{aligned} \mathbb{D} \left\{ \sum_{i=0}^n \eta(t_i)(w(t_{i+1}) - w(t_i)) \right\} &= \mathbb{E} \left\{ \sum_{i=0}^n \eta(t_i)(w(t_{i+1}) - w(t_i)) \right\}^2 = \\ &= \sum_{i=0}^n \sum_{j=0}^n \mathbb{E} [\eta(t_i)(w(t_{i+1}) - w(t_i))\eta(t_j)(w(t_{j+1}) - w(t_j))]. \end{aligned}$$

А двойная сумма разбивается на слагаемые двух видов: $i = j$ и $i \neq j$. Если $i = j$, то ввиду независимости приращений от прошлого

$$\begin{aligned} \mathbb{E} \{ \eta^2(t_i)(w(t_{i+1}) - w(t_i))^2 \} &= \mathbb{E} \{ \eta^2(t_i) \} \mathbb{E} \{ (w(t_{i+1}) - w(t_i))^2 \} = \\ &= \mathbb{E} \{ \eta^2(t_i) \} (t_{i+1} - t_i). \end{aligned}$$

Пусть $i \neq j$, для определенности мы будем считать, что $i < j$. Ввиду независимости приращений от прошлого имеем

$$\begin{aligned} \mathbb{E} \{ (\eta(t_i)(w(t_{i+1}) - w(t_i)))(\eta(t_j)(w(t_{j+1}) - w(t_j))) \} &= \\ &= \mathbb{E} \{ [(\eta(t_i)(w(t_{i+1}) - w(t_i))\eta(t_j)] (w(t_{j+1}) - w(t_j)) \} = \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E}[(\eta(t_i)(w(t_{i+1}) - w(t_i))\eta(t_j)] \mathbb{E}[w(t_{j+1}) - w(t_j)] = \\
&= \mathbb{E}[(\eta(t_i)(w(t_{i+1}) - w(t_i))\eta(t_j)] \times 0 = 0.
\end{aligned}$$

Окончательно мы получаем, что

$$\mathbb{D} \left\{ \sum_{i=0}^n \eta(t_i)(w(t_{i+1}) - w(t_i)) \right\} = \sum_{i=0}^n \mathbb{E} \{ \eta^2(t_i) \} (t_{i+1} - t_i).$$

Эта сумма очень похожа на интегральную сумму, но от какого интеграла? Этот интеграл легко записывается, и мы получаем более приятную формулу:

$$\mathbb{D} \left\{ \int_0^T \eta(t) dw(t) \right\} = \int_0^T \mathbb{E}[\eta(t)]^2 dt.$$

Из этой формулы, которая верна и для интегральных сумм, то есть для интегралов от кусочно-постоянных процессов $\eta(t)$, выводится и сходимость интегральных сумм к интегралу в случае, например, непрерывного процесса $\eta(t)$, причем непрерывность мы понимаем в среднем квадратическом, то есть

$$t_n \rightarrow t \Rightarrow \mathbb{E}[\eta(t_n) - \eta(t)]^2 \rightarrow 0.$$

Действительно, тогда по знаменитой теореме Кантора (функция, непрерывная на компакте, равномерно непрерывна) случайный процесс $\eta(t)$ равномерно непрерывен в среднем квадратическом, то есть по любому $\varepsilon > 0$ существует $\delta > 0$ такое, что

$$|t - s| < \delta \Rightarrow \mathbb{E}[\eta(t) - \eta(s)]^2 < \varepsilon.$$

Используя это и выбирая разбиение

$$\Delta : 0 = t_0 < t_1 < \dots < t_{n-1} < t_n = T$$

диаметра меньше δ , мы построим интегральную сумму

$$\sum_{i=0}^n \eta(t_i)(w(t_{i+1}) - w(t_i)).$$

Запишем эту сумму как интеграл от кусочно постоянной функции

$$\eta_\Delta(t) = \sum_{i=0}^{n-1} I_{[t_i, t_{i+1})}(t) \eta(t_i).$$

Рассмотрим также другое разбиение $\tilde{\Delta}$ с такой же оценкой для диаметра. Тогда мы имеем

$$\mathbb{E}[\eta_\Delta(t) - \eta(t)]^2 < \varepsilon,$$

$$\mathbb{E}[\eta_{\tilde{\Delta}}(t) - \eta(t)]^2 < \varepsilon.$$

Отсюда по тривиальному неравенству

$$(x - y)^2 \leq 2(x - z)^2 + 2(y - z)^2$$

мы имеем

$$\mathbb{E}[\eta_\Delta(t) - \eta_{\tilde{\Delta}}(t)]^2 < 4\varepsilon,$$

$$\mathbb{E} \left[\int_0^T \eta_\Delta(t) dw(t) - \int_0^T \eta_{\tilde{\Delta}}(t) dw(t) \right]^2 =$$

$$\mathbb{E} \left[\int_0^T \{\eta_\Delta(t) - \eta_{\tilde{\Delta}}(t)\} dw(t) \right]^2 = \int_0^T \mathbb{E} \{\eta_\Delta(t) - \eta_{\tilde{\Delta}}(t)\}^2 dt \leq 4\varepsilon \cdot T.$$

В последнем равенстве мы использовали формулу для дисперсии интеграла от кусочно постоянного процесса $\eta_\Delta(t) - \eta_{\tilde{\Delta}}(t)$. Резюмируем: если диаметр последовательности разбиений сходится к нулю, то соответствующие интегральные суммы сходятся и их предел называется интегралом Ито.

В интернете имеется курс лекций (на английском), который можно рекомендовать. Причем этот курс выкачивать не надо, если вы захотите его посмотреть, я дам файл, но не просто так, а потом проверю, что вы из него почерпнули.

Lawrence G. Evans. An Introduction to Stochastic Differential Equations. Version1.2. Berkeley University.

5. Стохастические дифференциальные уравнения

Теперь мы можем ввести понятие стохастического дифференциального уравнения. Для них неудобно использование обозначений типа $\frac{dw}{dt}$ (траектории винеровского процесса с вероятностью 1 недифференцируемы ни в одной точке), поэтому их предпочитают записывать в дифференциалах.

Решением X уравнения

$$dX = F(X)dt + G(X)dw(t)$$

по определению называется такой не зависящий от прошлого винеровского процесса (предсказуемый) случайный процесс X , что тождественно

$$X(t) = X(0) + \int_0^t F(X(s))ds + \int_0^t G(X(s))dw(s).$$

Таким образом, хотя стохастические дифференциальные уравнения записываются в дифференциалах, они на деле являются интегральными уравнениями, причем с использованием стохастических интегралов. Иногда они имеют явное решение, но как правило, для них разрабатывается техника приближенных вычислений. В теории обычных интегралов и дифференциальных уравнений используют приемы, которые называются 'замена переменных' и 'интегрирование по частям'.

6. Формула Ито замены переменных

Пусть X — предсказуемый процесс, причем

$$dX = Fdt + Gdw, \quad (4)$$

где F и G — некоторые предсказуемые процессы.

Вопрос. Чему равен стохастический дифференциал процесса

$$Y(t) = u(X(t), t)?$$

Ответ дается знаменитой формулой Ито:

$$dY(t) = \frac{\partial u}{\partial t}(X, t)dt + \frac{\partial u}{\partial x}(X, t)dX + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}(X, t)G^2(t)dt.$$

Более подробно можно записать следующим образом:

$$dY(t) = \frac{\partial u}{\partial t}(X, t)dt + \frac{\partial u}{\partial x}(X, t)F(t)dt + \frac{\partial u}{\partial x}(X, t)G(t)dw + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}(X, t)G^2(t)dt.$$

Доказательство этой формулы довольно длинно и технично, оно будет проведено на уровне идей. Поэтому нет смысла приводить также необходимые для справедливости формулы условия, их довольно много и все это можно найти в учебниках по теории стохастических дифференциальных уравнений.

Для понимания этой формулы рассмотрим частный случай функции u одной переменной, зависящей только от x и сравним формулу

$$dY(t) = \frac{\partial u}{\partial x}(X, t)dX + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}(X, t)G^2(t)dt.$$

в этой случае с обычной формулой замены переменной в интеграле. Мы видим, что в обычной формуле

$$\int u'(X)dX$$

отсутствует вторая производная

$$\frac{1}{2} \frac{\partial^2 u}{\partial x^2}(X, t)G^2(t)dt.$$

Почему в формуле Ито используется вторая и в то же время отсутствуют производные более высокого порядка?

Итак, как получается формула Ито. Мы можем приближенно написать приращение процесса $X(t)$:

$$X(t + \Delta t) - X(t) = F(t)\Delta t + G(t)[w(t + \Delta t) - w(t)].$$

Теперь посмотрим на приращение функции $u(X)$ и разложим его по формуле Тейлора.

$$\begin{aligned} u(X(t + \Delta t)) - u(X(t)) &\approx u'(X(t))(X(t + \Delta t) - X(t)) + \\ &+ \frac{1}{2}u''(X(t))(X(t + \Delta t) - X(t))^2 = \\ &= u'(X(t))(F(t)\Delta t + G(t)[w(t + \Delta t) - w(t)]) + \\ &+ \frac{1}{2}u''(X(t))(F(t)\Delta t + G(t)[w(t + \Delta t) - w(t)])^2 = \\ &= u'(X(t))(F(t)\Delta t + G(t)[w(t + \Delta t) - w(t)]) + \frac{1}{2}u''(X(t)) \times \\ &\times (F^2(t)\Delta t^2 + 2F(t)G(t)\Delta t[w(t + \Delta t) - w(t)] + G^2(t)[w(t + \Delta t) - w(t)]^2). \end{aligned}$$

Подумаем, какие слагаемые в

$$(F^2(t)\Delta t^2 + 2F(t)G(t)\Delta t[w(t + \Delta t) - w(t)] + G^2(t)[w(t + \Delta t) - w(t)]^2)$$

вносят существенный вклад после суммирования по всем приращениям.

Как мы уже неоднократно видели (например, в §14), сумма приращения $F^2(t)\Delta t^2 = o(\Delta t)$ мажорируется диаметром разбиения (умноженным на максимум F^2), поэтому стремится к нулю. В то же время, сумма слагаемых $G^2(t)[w(t + \Delta t) - w(t)]^2$ имеет большее значение. Мы уже считали

$$\mathbb{E}[w(t + \Delta t) - w(t)]^2 = \Delta t,$$

в то же время второй момент этого выражения равен

$$\mathbb{E}\{[w(t + \Delta t) - w(t)]^2\}^2 = 2\Delta t^2.$$

Таким образом, ошибка при замене в сумме случайной величины $[w(t + \Delta t) - w(t)]^2$ на его среднее Δt мажорируется диаметром разбиения, который сходится к нулю. Мы уже видели, что суммы всех таких приращений на отрезке $[0, T]$ сходятся при диаметре разбиения, сходящемся к 0, к T . Итак, слагаемое

$$G^2(t)[w(t + \Delta t) - w(t)]^2$$

нужно учитывать, заменив его на $G^2(t)\Delta t$. Рассмотрим другое слагаемое — $2F(t)G(t)\Delta t[w(t + \Delta t) - w(t)]$.

$$\mathbb{E}(\Delta t[w(t + \Delta t) - w(t)]) = 0, \quad \mathbb{E}(\Delta t[w(t + \Delta t) - w(t)])^2 = \Delta t^3.$$

Эти величины также являются маленькими. Мы увидели причину появления дополнительного слагаемого. Напоминаю, что все это нестрого и не может считаться настоящим доказательством формулы Ито.

7. Применения формулы Ито.

Пример 1. Теперь покажем, как с помощью формулы Ито мы можем опять сосчитать интеграл

$$\int_0^T w(t)dw(t).$$

Согласно формуле Ито

$$dw^2(t) = 2w(t)dw(t) + \frac{1}{2}2dt.$$

Переходя к интегралам

$$\begin{aligned} \int_0^T dw^2(t) &= 2 \int_0^T w(t)dw(t) + \int_0^T dt, \\ w^2(T) &= 2 \int_0^T w(t)dw(t) + T, \end{aligned}$$

что и требовалось.

Замечание. Формула Ито имеет смысл и для функций многих переменных от нескольких случайных процессов. Используя такую формулу, легко получить следующую:

Формула Ито для произведения

Пусть

$$dX_1 = F_1 dt + G_1 dw,$$

$$dX_2 = F_2 dt + G_2 dw,$$

тогда

$$d(X_1 X_2) = X_2 dX_1 + X_1 dX_2 + G_1 G_2 dt.$$

Однако эту формулу можно вывести и из трехкратного использования одномерной формулы Ито, используя равенство

$$d(X_1 X_2) = \frac{1}{2} \{ d(X_1 + X_2)^2 - d(X_1)^2 - d(X_2)^2 \}.$$

8. Пример решения стохастического дифференциального уравнения

$$dX(t) = g(t)X(t)dw(t),$$

где g – непрерывная (детерминированная, то есть неслучайная) функция, $X(0) = 1$.

Тогда

$$X(t) = e^{-\frac{1}{2} \int_0^t g^2(s)ds + \int_0^t gdw}.$$

Обозначим

$$Y(t) = -\frac{1}{2} \int_0^t g^2(s)ds + \int_0^t gdw,$$

тогда

$$dY(t) = -\frac{1}{2}g^2(t)dt + g(t)dw(t).$$

По формуле Ито при $X(t) = u(Y) = e^Y$

$$\begin{aligned} dX &= \frac{\partial u}{\partial x}(Y)dY + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(Y)g^2dt = \\ &= e^Y \left\{ -\frac{1}{2}g^2(t)dt + g(t)dw(t) + \frac{1}{2}g^2dt \right\} = gXdw. \end{aligned}$$

§35. Мартингалы

В основе понятия *мартингала* лежит другое важнейшее понятие — *условное среднее* или *условное математическое ожидание*. А именно, пусть дана случайная величина ξ на вероятностном пространстве $(\Omega, \mathcal{A}, \mathbb{P})$ (т.е. измеримая относительно σ -алгебры \mathcal{A}), имеющая среднее значение $\mathbb{E}\xi$, и пусть дана σ -подалгебра $\mathfrak{B} \subset \mathcal{A}$. Тогда существует такая \mathfrak{B} -измеримая случайная величина $\mathbb{E}(\xi|\mathfrak{B})$, связанная с ξ соотношением:

$$\int_B \xi d\mathbb{P} = \int_B \mathbb{E}(\xi|\mathfrak{B}) d\mathbb{P} \text{ для всех } B \in \mathfrak{B}. \quad (1)$$

Существование такой функции основано на использовании теоремы Радона – Никодима. Согласно этой теореме, для любого заряда λ на σ -алгебре \mathcal{A} (т. е. счетно аддитивного отображения в \mathbf{R}), абсолютно непрерывного относительно вероятности \mathbb{P} , т. е. такого, что

$$\mathbb{P}(A) = 0 \Leftrightarrow \lambda(A) = 0.$$

В этой ситуации теорема Радона – Никодима утверждает, что существует так называемая производная Радона – Никодима $f = \frac{d\lambda}{d\mathbb{P}}$, которая превращает заряд λ в неопределенный интеграл по f :

$$\lambda(A) = \int_A f(\omega) d\mathbb{P}(\omega) \text{ для всех } A \in \mathcal{A}.$$

Как же из этой теоремы мы выводим существование условного среднего значения? Мы вводим заряд λ на \mathfrak{B} как неопределенный интеграл:

$$\lambda(B) = \int_B \xi d\mathbb{P}, \text{ для всех } B \in \mathfrak{B}.$$

Из свойств интеграла Лебега следует абсолютная непрерывность λ относительно \mathbb{P} . Условным средним мы называем производную Радона – Никодима заряда λ . Таким образом, мы сначала 'превращаем' случайную величину ξ в заряд λ , сужаем область определения заряда на \mathfrak{B} , а потом 'превращаем' суженный заряд в случайную величину.

Лучше всего можно понять условное среднее в случае конечной σ -алгебры \mathfrak{B} , порожденной конечным числом событий B_1, \dots, B_k , этот случай был по-существу разобран в разделе 'Функция регрессии'. Нетрудно видеть, что в этой σ -алгебре имеются наименьшие непустые элементы $A_1 + \dots + A_n = \Omega$ и каждый элемент σ -алгебры представляется в виде конечной суммы этих элементов. Как выглядит в этой ситуации обычная условная вероятность некоторого события B , которую мы представляем как условное среднее случайной величины I_B ? Теперь это не число, а случайная величина, но она записывается через условные вероятности:

$$\mathbb{E}(I_B|\mathfrak{B})(\omega) = \mathbb{P}(B|A_i), \text{ если } \omega \in A_i,$$

эта случайная величина задана с точностью до множества меры 0.

Упражнение 4. Пусть ξ — случайная величина, имеющая среднее и измерима относительно σ -алгебры \mathfrak{B} . Покажите, что

$$\mathbb{E}(\xi|\mathfrak{B}) = \xi.$$

1. Свойства условного среднего.

Условное среднее ведет себя, как усреднение интеграла (но на каждом A_i в рассмотренном примере усреднение идет только по A_i).

Например, легко видеть, что

$$\mathbb{E}(a\xi + b\eta|\mathfrak{B}) = a\mathbb{E}(\xi|\mathfrak{B}) + b\mathbb{E}(\eta|\mathfrak{B}).$$

Отметим наиболее важные свойства:

- i) $\mathbb{E}(\mathbb{E}(\xi|\mathfrak{B})) = \mathbb{E}(\xi)$,
- ii) если случайная величина η измерима относительно σ -алгебры \mathfrak{B} и ограничена, а ξ интегрируема, то

$$\mathbb{E}(\xi\eta|\mathfrak{B}) = \eta\mathbb{E}(\xi|\mathfrak{B}).$$

Это утверждение немедленно доказывается в случае $\eta = I_B$ — индикатора события $B \in \mathfrak{B}$, а именно, для $I_B\mathbb{E}(\xi|\mathfrak{B})$ проверяется (1). Произвольный случай сводится к этому: сначала мы доказываем ii) для дискретных η , а потом предельным переходом для произвольных η . (Рекомендуется самостоятельно провести эти рассуждения.)

- iii) если φ — выпуклая функция, т. е. для любого $\lambda \in (0, 1)$

$$\varphi(\lambda x + (1 - \lambda)y) \leq \lambda\varphi(x) + (1 - \lambda)\varphi(y),$$

то

$$\varphi(\mathbb{E}(\xi|\mathfrak{B})) \leq \mathbb{E}(\varphi(\xi)|\mathfrak{B}). \quad (*)$$

Сначала нужно доказать неравенство для обычного среднего

$$\varphi(\mathbb{E}(\xi)) \leq \mathbb{E}(\varphi(\xi)). \quad (**)$$

Доказательство проводится сначала для дискретной случайной величины с конечным числом значений, при этом все сводится к неравенству, доказываемому по индукции: если $p_1 + \dots + p_n = 1$, $p_i > 0$, то

$$\varphi(p_1x_1 + \dots + p_nx_n) \leq p_1\varphi(x_1) + \dots + p_n\varphi(x_n).$$

Дальше предельным переходом мы доказываем (**) для любой дискретной случайной величины, а потом и для любой случайной величины (у которой существует соответствующие средние). Неравенство для условного среднего сводится к неравенству для обычного среднего: сначала мы рассматриваем случай конечной σ -алгебры \mathfrak{B} , потом предельным переходом рассматриваем произвольный случай. При этом конечные σ -алгебры мы строим по случайной величине $\mathbb{E}(\xi|\mathfrak{B})$. (Имеются в виду σ -алгебры, порожденные событиями $\{\xi \in [\frac{k}{n}, \frac{k+1}{n}] \}$, $-N \leq k \leq N-1$, $\{\xi \in [-\infty, \frac{-N}{n}] \}$, $\{\xi \in [\frac{N}{n}, \infty) \}$).

А теперь перейдем к мартингалам.

2. Мартингалы.

Определение. Пусть для каждого $t \geq 0$ задана σ -алгебра $\mathfrak{F}_t \subset \mathfrak{A}$, причем $\mathfrak{F}_t \subset \mathfrak{F}_s$ при $t \leq s$. Такой набор σ -алгебр называется *фильтрацией*. Случайный процесс ξ_t называется мартингалом относительно этой фильтрации, если при любом t случайная величина ξ_t измерима относительно \mathfrak{F}_t и при любых $t \leq s$

$$\mathbb{E}(\xi_s|\mathfrak{F}_t) = \xi_t.$$

Это определение аналогичным образом формулируется и для дискретного времени $\{0\} \cup \mathbf{N}$.

Рассмотрим простой пример, в котором Ω состоит из всех бесконечных последовательностей из гербов и решек, \mathfrak{F}_1 состоит из Ω , \emptyset , множества всех последовательностей вида $\{\text{г...}\}$ (вероятность этого множества равна $1/2$), и множества всех последовательностей вида $\{\text{р...}\}$ (вероятность этого множества также равна $1/2$). Соответственно \mathfrak{F}_2 порождается множествами $\{\text{гг...}\}$, $\{\text{гр...}\}$, $\{\text{рг...}\}$, $\{\text{рр...}\}$ (с вероятностью $1/4$ каждое). Аналогично строятся и все \mathfrak{F}_n для

больших n . Пусть случайные величины ξ_n равны 1 если на n -м месте в элементе находится герб и -1 если на n -м месте находится решка. Эти случайные величины измеримы относительно σ -алгебры \mathfrak{F}_n и $|\xi_n| \leq 1$. Введем также случайную величину $\eta = \sum_n \xi_n / 2^n$. Очевидно, что случайная величина η интегрируема. Определим последовательность

$$\eta_n = \mathbb{E}(\eta | \mathfrak{F}_n)$$

Легко проверяется, что последовательность (η_n) представляет собой мартингал, который определяется фильтрацией и одной случайной величиной η . Более того,

$$\eta_n = \sum_{i=1}^n \xi_i / 2^i \xrightarrow{\text{П.Н.}} \eta.$$

Ниже мы докажем, что при не очень ограничительных условиях любой мартингал имеет подобное свойство.

3. Теорема Дуба. *Пусть последовательность (η_n) является мартингалом относительно фильтрации (\mathfrak{F}_n) , причем $\sup_n \mathbb{E}|\eta_n| < \infty$. Тогда последовательность (η_n) сходится п. н.*

Сначала докажем неравенство для мартингалов. Предварительно напомним обозначение для любой случайной величины $\xi^+ = \max\{\xi, 0\}$.

Лемма. *В условиях теоремы зафиксируем набор η_1, \dots, η_N и зафиксируем числа $a < b$. Обозначим через β_N случайную величину — число всех перескоков последовательности (η_n) через отрезок $[a, b]$, то есть число всех кусков в нашем наборе вида $\eta_i(\omega), \eta_{i+1}(\omega), \dots, \eta_j(\omega)$, где $i < j$, и $\eta_i(\omega) \leq a, \eta_j(\omega) \geq b, \eta_k(\omega) < b$ для всех $i < k < j$. Тогда*

$$\mathbb{E}\beta_N \leq \frac{\mathbb{E}(\eta_N - a)^+}{b - a}. \quad (2)$$

Доказательство. Обозначим $\xi_i = (\eta_i - a)^+$. Заметим, что в терминах новых случайных величин ξ_i один перескок означает изменение от $\xi_i = 0$ до значения $\xi_j \geq b - a$, где $i < j$ и $\xi_k < b - a$ для всех $i < k < j$, т.е. момент j — момент первого перескока после i . Далее обозначим через A_k событие, состоящее из таких ω , что $\xi_i(\omega) = 0$ для некоторого $i < k$ и $\xi_m(\omega) < b - a$ для всех $i \leq m < k$. Событие A_k отнюдь не означает завершение перескока в момент k . Но если для данного случайного исхода ω происходит (один!) перескок от $\xi_i(\omega) = 0$ до $\xi_j(\omega) \geq b - a$, то ω попадает во все A_k , где $i < k \leq j$. Поэтому для всех ω , в которых начиная с момента i и кончая моментом j произошел перескок, имеет место неравенство

$$b - a \leq \sum_{k=i}^{j-1} (\xi_{k+1} - \xi_k) I_{A_{k+1}}.$$

Суммируя эти неравенства для всех ω и всех перескоков, получаем

$$(b - a)\beta_N \leq \sum_{i=1}^{N-1} (\xi_{i+1} - \xi_i) I_{A_{i+1}}. \quad (3)$$

Заметим, что к концу нашей последовательности может появиться $\xi_i(\omega) = 0$, но после этого никакого перескока не будет. При этом ω все равно попадет во все A_k , где $k > i$. Тогда левая часть (3) не увеличится за счет соответствующих слагаемых, но и правая часть не уменьшится, так как

$$\sum_{k=i+1}^N \xi_k(\omega) - \xi_{k-1}(\omega) = \xi_N - \xi_i \geq 0.$$

Упражнение. Дан конечный набор вещественных чисел $\{x_1, \dots, x_n\}$. Выберем в нем все числа x_i такие, что для данного i существует $j(i) \geq i$ со свойством $x_i + \dots + x_{j(i)} \geq 0$. Доказать, что сумма всех выбранных чисел неотрицательна.

Теперь заметим, что функция $\varphi : x \rightarrow \max\{x - a, 0\}$ является выпуклой функцией, $\xi_i = \varphi(\eta_i)$. По свойству iii)

$$\mathbb{E}(\xi_i | \mathfrak{F}_i) = \xi_i = \varphi(\eta_i) = \varphi(\mathbb{E}(\eta_{i+1} | \mathfrak{F}_i)) \leq \mathbb{E}(\varphi(\eta_{i+1}) | \mathfrak{F}_i) = \mathbb{E}(\xi_{i+1} | \mathfrak{F}_i) \text{ для любого } i.$$

Существенно, что входящие в (3) индикаторы $I_{A_{i+1}}$ зависят лишь от значений мартингала до момента i включительно и поэтому принадлежат \mathfrak{F}_i . Теперь выведем (2) из (3), используя свойства i), ii), iii) условного среднего:

$$\begin{aligned} (b-a)\mathbb{E}\beta_N &\leq \mathbb{E} \left[\sum_{i=1}^{N-1} (\xi_{i+1} - \xi_i) I_{A_{i+1}} \right] = \\ \sum_{i=1}^{N-1} \mathbb{E} [(\xi_{i+1} - \xi_i) I_{A_{i+1}}] &= \sum_{i=1}^{N-1} \mathbb{E} [\mathbb{E}((\xi_{i+1} - \xi_i) I_{A_{i+1}} | \mathfrak{F}_i)] = \\ (\text{пока мы использовали i})) &= \sum_{i=1}^{N-1} \mathbb{E} [I_{A_{i+1}} \mathbb{E}(\xi_{i+1} - \xi_i | \mathfrak{F}_i)] \end{aligned}$$

(здесь мы использовали свойство ii); как мы уже видели раньше, из iii) следует, что $\mathbb{E}(\xi_{i+1} - \xi_i | \mathfrak{F}_i) \geq 0$, поэтому заменяя $I_{A_{i+1}}$ на 1, мы только усилим неравенство)

$$\leq \sum_{i=1}^{N-1} \mathbb{E}[\mathbb{E}(\xi_{i+1} - \xi_i | \mathfrak{F}_i)] =$$

(а теперь используем i) в другую сторону)

$$= \sum_{i=1}^{N-1} \mathbb{E}(\xi_{i+1} - \xi_i) = \mathbb{E}\xi_N - \mathbb{E}\xi_1 \leq \mathbb{E}\xi_N.$$

(последнее неравенство следует из неотрицательности ξ_1).

Доказательство теоремы. Предположим противное, предел $(\eta_n)(\omega)$ не существует для ω на множестве A строго положительной меры (убедитесь сами, что это множество измеримо). Это означает

$$\liminf_n \eta_n(\omega) < \limsup_n \eta_n(\omega) \text{ для всех } \omega \in A.$$

Это можно записать и по другому:

$$\liminf_n \eta_n(\omega) < a(\omega) < b(\omega) < \limsup_n \eta_n(\omega) \text{ для всех } \omega \in A.$$

Разумеется, с каждыми $a < b$ мы можем связать событие

$$A_{a,b} = \{\omega : \liminf_n \eta_n(\omega) < a < b < \limsup_n \eta_n(\omega)\}, A = \bigcup_{a < b} A_{a,b}.$$

Последнее равенство верно, если мы рассматриваем в нем лишь пары рациональных a и b , а множество таких пар счетно. Далее, объединение счетного числа множеств нулевой меры имеет меру нуль, поэтому из $\mathbb{P}A > 0$ следует существование такой пары рациональных чисел $a < b$, что $\mathbb{P}A_{a,b} > 0$. Однако по определению $A_{a,b}$ для любого $\omega \in A_{a,b}$ последовательность $(\eta_n(\omega))$ имеет бесконечное число пересеков через (a, b) . Поэтому

$$\lim_N \mathbb{E}\beta_N = \infty \mathbb{P}A_{a,b} = \infty.$$

Вместе с неравенством (2) это приводит к противоречию с ограниченностью последовательности $\mathbb{E}|\eta_n|$.

Пример мартингала, для которого не выполняются условия теоремы Дуба. Рассмотрим процесс случайного блуждания ξ_n .

4. Некоторые вопросы.

Что такое момент остановки, если задана возрастающая фильтрация σ -алгебр \mathfrak{F}_t , $t \geq 0$. Это такая неотрицательная величина τ , что

$$\{\tau \leq t\} \subset \mathfrak{F}_t \text{ для всех } t.$$

Иными словами, для принятия решения об остановке до момента t включительно, нам достаточно наблюдать случайный процесс до этого момента включительно.

Упражнения. 1. Пусть S_1, \dots, S_n — мартингал с конечным временем, τ — случайный момент остановки. Доказать, что $\mathbb{E}S_\tau$ не зависит от τ .

2. Пусть ξ_t — случайный процесс, непрерывный в среднем, τ — некоторый момент остановки. Доказать, что ξ_τ — случайная величина, то есть измерима.

3. Как доказать, что два условных средних $\mathbb{E}(\xi|\mathfrak{F}_t)$ и $\mathbb{E}(\eta|\mathfrak{F}_t)$ совпадают почти наверное?

5. Винеровский процесс — мартингал.

Упражнение. Пусть случайная величина ξ имеет среднее и не зависит от σ -алгебры \mathfrak{F} . Тогда

$$\mathbb{E}(\xi|\mathfrak{F}) = \mathbb{E}\xi \text{ п. н.}$$

(Нетрудно догадаться, что означает независимость от σ -алгебры: для любого boreлевского множества B и любого события $A \in \mathfrak{F}$ события $\{\xi \in B\}$ и A должны быть независимы). Как и обычное среднее, условное среднее также удовлетворяет неравенству $|\mathbb{E}(\xi|\mathfrak{F})| \leq Const$, если $|\xi| \leq Const$ п. н. Поэтому, представляя случайную величину как равномерный предел последовательности дискретных случайных величин, как обычно, мы сводим общий случай к дискретному случаю, а дискретный случай ввиду аддитивности условного среднего сводится к случаю индикатора события. Итак, событие A не зависит от всех $B \in \mathfrak{F}$. Проверьте, что постоянная функция $\mathbb{P}(A)$ удовлетворяет определению $\mathbb{E}(I_A|\mathfrak{F})$.

Из этого упражнения легко следует, что винеровский процесс $w(t)$ является мартингалом относительно фильтрации

$$\mathfrak{F}_t = \text{наименьшая } \sigma\text{-алгебра, относительно которой измеримы все } w(s), s \leq t.$$

Проверьте, что если $u \geq t$, то

$$\mathbb{E}(w(t)|\mathfrak{F}_t) = w(t), \quad \mathbb{E}(w(u) - w(t)|\mathfrak{F}_t) = \mathbb{E}(w(u) - w(t)) = 0.$$

Из этого немедленно следует, что w — мартингал.

§36. Задача о разборчивой невесте

Одним из примеров использования понятия мартингала является 'Задача о разборчивой невесте' (другое название — 'Задача о выборе секретаря'). Рассматривается следующая ситуация — невеста выбирает себе мужа из N женихов. Число женихов известно заранее. Невеста может лишь сравнивать женихов между собой, и обладает абсолютной памятью, то есть может сравнивать данного жениха со всеми виденными ею ранее. Для любых двух женихов она может сказать, кто из них для нее лучше и кто — хуже, причем эта ранжировка имеет все свойства порядка — если A лучше B , B лучше C , то A лучше C . Существенное и жесткое условие — отвергнутый жених перестает участвовать в конкурсе, передумать по отношению к нему уже нельзя. Думается, что первое название лучше соответствует задаче, подчеркивая ее условный сказочный характер.

1. Постановка задачи.

Цель невесты в данной задаче — сделать наибольшей вероятность выбора наилучшего жениха среди всех N . Заметим, что в данной ситуации возможны и более прагматичные цели — например, если все женихи упорядочены по качеству — 1-й (наилучший), 2, 3, ..., N -й (наихудший), то можно минимизировать среднее значение номера выбранного жениха. Однако задачей в такой постановке мы заниматься не будем. (Более того, непонятно обоснование такой постановки задачи.)

Единственный приходящий в голову алгоритм — отказ от выбора до n -го жениха, после чего нужно выбрать первого, который будет лучше всех предыдущих. Наша последующая задача — оптимизация выбора номера n , а также доказательство (интуитивно понятного) утверждения, что все другие алгоритмы хуже.

2. Математическая модель.

Если мы говорим о вероятности события и о среднем значении случайной величины, то мы должны иметь также пространство элементарных исходов с заданными на них вероятностями. В данной задаче такая модель строится совершенно очевидным способом. Произвольный элементарный исход ω — это порядок прохождения женихов перед невестой. Как вы знаете, множество из N элементов можно упорядочить $N!$ способами, таким образом, $|\Omega| = N!$. Итак, $\omega = (\omega_1, \dots, \omega_N)$ — некоторая перестановка чисел $\{1, 2, \dots, N\}$, ранжировка женихов невестой уже после знакомства со всеми. При этом σ -алгеброй событий является σ -алгебра \mathfrak{A} всех подмножеств Ω . Минимальными непустыми элементами этой σ -алгебры являются элементарные исходы — точки Ω .

Сама постановка задачи определяет естественную возрастающую фильтрацию σ -подалгебр \mathfrak{A} : Впрочем, здесь есть над чем подумать. Удобнее всего определять конечные σ -подалгебры заданием их минимальных элементов (своего рода элементарных исходов для данной σ -подалгебры). Все остальные элементы σ -подалгебры будут объединениями минимальных элементов. Направляется следующая фильтрация: минимальный элемент n -й σ -подалгебры должен определяться числами $(\omega_1, \dots, \omega_n)$ и в него должны входить все элементарные исходы $\omega = (\omega_1, \dots, \omega_n)$, первые n чисед в которых совпадают с первыми n числами набора ω . Нетрудно видеть, что число элементарных исходов в такой σ -подалгебре будет равно $N!/(N - n)!$ (так как число способов переупорядочить числа, не попавшие в $(\omega_1, \dots, \omega_n)$, равно $(N - n)!$). Однако такая фильтрация не соответствует условиям задачи, так как числа $\omega_1, \dots, \omega_n$ становятся известными лишь в случае знакомства со всеми женихами. При построении фильтрации надо исходить из ситуации.

Итак, каждый минимальный элемент \mathfrak{A}_n определяется известным на n -й момент упорядочением первых n женихов. Это упорядочение мы запишем как $(\sigma_1, \dots, \sigma_n)$ — это некоторая перестановка чисел $\{1, 2, \dots, n\}$. В минимальный элемент \mathfrak{A}_n , задаваемый перестановкой $(\sigma_1, \dots, \sigma_n)$ мы включаем все элементарные исходы $(\omega_1, \dots, \omega_n) \in \Omega$, удовлетворяющих условиям: для всех пар натуральных чисел i, j

leqn неравенство $\sigma_i < \sigma_j$ выполняется тогда и только тогда, когда $\omega_i < \omega_j$. Итак, число минимальных элементов в \mathfrak{A}_n равно $n!$, каждый такой минимальный элемент содержит $N!/n!$ элементарных исходов из Ω .

3. Задача об оптимальном моменте остановки.

Сначала мы рассмотрим эту задачу в более общей ситуации.

На вероятностном пространстве $(\Omega, \mathfrak{A}, \mathbb{P})$ заданы возрастающая фильтрация σ -подалгебр $\{\mathfrak{A}_1, \mathfrak{A}_2, \dots, \mathfrak{A}_N = \mathfrak{A}\}$ и такой набор случайных величин $\{\xi_1, \dots, \xi_N\}$, что случайная величина ξ_i измерима относительно σ -подалгебры \mathfrak{A}_i для всех i . Рассмотрим также множество \mathbb{T} всех случайных моментов остановки τ со значениями в $\{1, \dots, N\}$. Напомним, что моментом остановки τ называется случайная величина со следующим свойством:

$$\{\tau \leq n\} \in \mathfrak{A}_n \quad \forall n \in \{1, \dots, N\}.$$

Для упрощения понимания ситуации мы будем считать Ω конечным (как и в нашей задаче). Тогда каждая \mathfrak{A}_i будет также конечной, а следовательно, будет порождаться конечным числом минимальных попарно несомненных элементов. Эти элементы целесообразно мыслить

элементарными исходами, которые получены к моменту времени i . Таким образом, в каждый момент времени мы имеем множество Ω_i элементарных исходов, каждый из которых является некоторым подмножеством (событием) в момент времени $i+1$. Определение момента остановки означает, что в каждый момент времени i полученный к этому моменту элементарный исход $\omega^{(i)}$ дает достаточную информацию для проверки выполнения или невыполнения равенства $\tau = i$.

Итак, наша задача — найти момент остановки, на котором достигается максимум

$$V = \sup \{\mathbb{E}\xi_\tau : \tau \in \mathbb{T}\}.$$

Задача решается поэтапно, с помощью индуктивного процесса, причем индукция проводится не снизу, с $n = 1$, а сверху, с $n = N$. Введем множество \mathbb{T}_n всех элементов \mathbb{T} , принимающих значения в $\{n, \dots, N\}$. Очевидно, что $\mathbb{T}_1 = \mathbb{T}$.

$$V_n = \sup \{\mathbb{E}\xi_\tau : \tau \in \mathbb{T}_n\}.$$

\mathbb{T}_N состоит лишь из одного элемента $\tau \equiv N$. Легко определяется максимизирующий τ_{N-1} во множестве \mathbb{T}_{N-1} . Действительно,

$$\tau_{N-1}(\omega) = \begin{cases} N, & \text{если } \xi_{N-1}(\omega) < \mathbb{E}(\xi_N | \mathfrak{A}_{N-1})(\omega), \\ N-1 & \text{если } \xi_{N-1}(\omega) \geq \mathbb{E}(\xi_N | \mathfrak{A}_{N-1})(\omega). \end{cases}$$

Далее мы введем по индукции новые случайные величины η_n , заданные соотношениями $\eta_N = \xi_N$, $\eta_n = \max\{\xi_n, \mathbb{E}(\eta_{n+1} | \mathfrak{A}_n)\}$. Далее

$$\tau_n(\omega) = \inf\{i \geq n : \xi_i(\omega) = \eta_i(\omega)\}. \quad (1)$$

При $n = 1$ эта конструкция дает решение общей задачи, но нам нужно кое-что доказать.

1) Сначала мы должны показать, что τ_n является моментом остановки. Тогда на каждом шаге мы сможем остановиться, исходя лишь из имеющейся на данный момент информации. Но это очевидно, для $\tau_n = i$ необходимо и достаточно, чтобы для всех $j < i$ выполнялось $\xi_i(\omega) < \eta_i(\omega)$, а для i было бы равенство $\xi_i(\omega) = \eta_i(\omega)$.

2) $\mathbb{E}(\xi_{\tau_n} | \mathfrak{A}_n) = \eta_n \geq \mathbb{E}(\xi_\tau | \mathfrak{A}_n)$ для любого $\tau \in \mathbb{T}_n$.

Интегрирование этого неравенства по Ω устанавливает оптимальность момента остановки τ_n .

Итак, доказываем 2). Напомним, что доказательство неравенства

$$\mathbb{E}(\xi | \mathfrak{F}) \geq \mathbb{E}(\eta | \mathfrak{F})$$

для почти всех ω означает доказательство неравенства

$$\int_A \mathbb{E}(\xi | \mathfrak{F}) d\mathbb{P} \geq \int_A \mathbb{E}(\eta | \mathfrak{F}) d\mathbb{P}$$

для всех $A \in \mathfrak{F}$. Мы знаем также, что если $A \in \mathfrak{F}$, то

$$\int_A \xi d\mathbb{P} = \int_A \mathbb{E}(\xi | \mathfrak{F}) d\mathbb{P}.$$

Начинаем доказательство. По индукции мы предполагаем, что 2) выполняется для всех n больше или равных данного n и докажем свойство 2) для $n = n - 1$. Итак, $A \in \mathfrak{A}_{n-1}$, $\tau \in \mathbb{T}_{n-1}$. Мы также обозначим $\tau' = \max\{\tau, n\}$.

$$\int_A \xi_\tau d\mathbb{P} = \int_{A \cap \{\tau = n-1\}} \xi_\tau d\mathbb{P} + \int_{A \cap \{\tau \geq n\}} \xi_\tau d\mathbb{P} =$$

$$= \int_{A \cap \{\tau=n-1\}} \xi_{n-1} d\mathbb{P} + \int_{A \cap \{\tau \geq n\}} \xi_{\tau'} d\mathbb{P} =$$

(не только $A \in \mathfrak{A}_{n-1}$, но и $\{\tau \geq n\} \in \mathfrak{A}_{n-1}$ (кстати, почему?), поэтому во втором интеграле мы можем заменить подинтегральную случайную величину на ее условное математическое ожидание относительно \mathfrak{A}_{n-1})

$$= \int_{A \cap \{\tau=n-1\}} \xi_{n-1} d\mathbb{P} + \int_{A \cap \{\tau \geq n\}} \mathbb{E}(\mathbb{E}(\xi_{\tau'} | \mathfrak{A}_n) | \mathfrak{A}_{n-1}) d\mathbb{P} \leq$$

(а теперь мы используем предположение индукции для $n = n$ и для $\tau = \tau' \in \mathbb{T}_n$)

$$\leq \int_{A \cap \{\tau=n-1\}} \xi_{n-1} d\mathbb{P} + \int_{A \cap \{\tau \geq n\}} \mathbb{E}(\eta_n | \mathfrak{A}_{n-1}) d\mathbb{P} \leq$$

(далее мы применяем то, что η_{n-1} мажорирует как ξ_{n-1} , так и $\mathbb{E}(\eta_n | \mathfrak{A}_{n-1})$, к обоим интегралам)

$$\leq \int_A \eta_{n-1} d\mathbb{P}.$$

Теперь, используя индуктивное предположение $\xi_{\tau_n} = \eta_n$, а также совпадение по построению τ_{n-1} и τ_n на событии $\{\tau_{n-1} \geq n\}$ нужно проверить, что для $\tau = \tau_{n-1}$ наши неравенства являются равенствами. Итак, согласно предположению индукции

$$\begin{aligned} \int_{A \cap \{\tau \geq n\}} \mathbb{E}(\mathbb{E}(\xi_{\tau'_{n-1}} | \mathfrak{A}_n) | \mathfrak{A}_{n-1}) d\mathbb{P} &= \int_{A \cap \{\tau \geq n\}} \mathbb{E}(\mathbb{E}(\xi_{\tau_n} | \mathfrak{A}_n) | \mathfrak{A}_{n-1}) d\mathbb{P} = \\ &= \int_{A \cap \{\tau \geq n\}} \mathbb{E}(\eta_n | \mathfrak{A}_{n-1}) d\mathbb{P}. \end{aligned}$$

По построению

$$\eta_{n-1} = \begin{cases} \xi_{n-1} & \text{на } \{\tau_{n-1} = n-1\} \\ \mathbb{E}(\eta_n | \mathfrak{A}_{n-1}) & \text{на } \{\tau_{n-1} \geq n\}. \end{cases}$$

4. Решение задачи о разборчивой невесте

Теперь перейдем к конкретной задаче. Итак, Ω состоит из всех наборов $(\omega_1, \dots, \omega_N)$ различных перестановок чисел между 1 и N . $\omega_i = 1$ означает, что i -й жених является лучшим. Нам надо выбрать такой момент остановки τ , что вероятность $V = \mathbb{P}\{\omega_\tau = 1\}$ является максимальной. К сожалению, случайные величины ω_i не являются адаптированными, чтобы знать их значения, мы должны знать всю перестановку женихов. Но ситуацию спасает другое представление V :

$$V = \mathbb{E}(\xi_\tau),$$

где $\xi_n = \mathbb{P}\{\omega_n = 1 | \mathfrak{A}_n\}$. Заметим, что $\{\tau = n\} \in \mathfrak{A}_n$. Поэтому мы можем использовать для доказательства следующую цепочку равенств:

$$\begin{aligned} V &= \mathbb{P}\{\omega_\tau = 1\} = \sum_{n=1}^N \int_{\tau=n} I_{\{\omega_n=1\}} d\mathbb{P} = \\ &= \sum_{n=1}^N \int_{\tau=n} \mathbb{E}(I_{\{\omega_n=1\}} | \mathfrak{A}_n) d\mathbb{P} = \sum_{n=1}^N \int_{\tau=n} \xi_n d\mathbb{P} = \mathbb{E}\xi_\tau. \end{aligned}$$

Для анализа величин ξ_n нам будет удобно определить σ -алгебры \mathfrak{A}_n как наименьшие σ -алгебры, относительно которых измеримы для всех $m \leq n$ случайные величины φ_m , равные для данного $\omega = (\omega_1, \dots, \omega_m)$ числу всех ω_i , $i \leq m$, не худших ω_m . Действительно, легко видеть, что задание значений всех φ_m , $m \leq n$, однозначно определяет элементарный исход $(\omega_1, \dots, \omega_n)$ для σ -алгебры \mathfrak{A}_n .

Любопытно, что случайные величины φ_n независимы между собой. Действительно, $1 \leq \varphi_n \leq n$, $\mathbb{P}\{\varphi_n = i\} = 1/n$ для всех $i \leq n$,

$$\mathbb{P}\{\varphi_1 = i_1, \varphi_2 = i_2, \dots, \varphi_N = i_N\} = 1/N!.$$

Заметим, что случайная величина ξ_n является функцией φ_n . Действительно,

$$\xi_n = \begin{cases} n/N, & \text{если } \varphi_n = 1, \\ 0, & \text{если } \varphi_n > 1. \end{cases}$$

Все же объясним это равенство. Если $\varphi_n > 1$, то среди первых $n - 1$ чисел ω_i найдется по крайней мере одно меньшее ω_n . Это верно для всех ω из одного и того же минимального элемента \mathfrak{A}_n . Но тогда $I_{\{\omega_n=1\}}$ равен 0 на этом элементе. Тем более равно 0 условное среднее — усреднение $I_{\{\omega_n=1\}}$ по этому элементу. Теперь рассмотрим ситуацию $\varphi_n = 1$. В этом случае ω_n является лучшим среди первых n элементов набора ω . Но вероятность того, что лучший элемент набора ω находится среди первых n элементов ω равна n/N . Из нашей формулы следует, что величины ξ_n также независимы. Это облегчает процесс построения случайных величин η_n из предыдущего пункта. Мы имеем,

$$\eta_N = \xi_N = \begin{cases} 1, & \text{если } \varphi_N = 1, \\ 0, & \text{если } \varphi_N > 1. \end{cases}$$

Тогда в силу независимости ξ_N от \mathfrak{A}_{N-1} мы имеем

$$\begin{aligned} \mathbb{E}(\xi_N | \mathfrak{A}_{N-1}) &= 1/N, \\ \eta_{N-1} = \max\{\xi_{N-1}, \mathbb{E}(\eta_N | \mathfrak{A}_{N-1})\} &= \begin{cases} \frac{N-1}{N}, & \text{если } \varphi_{N-1} = 1 \\ 1/N, & \text{если } \varphi_{N-1} > 1, \end{cases} \\ \mathbb{E}(\eta_{N-1} | \mathfrak{A}_{N-2}) &= \frac{1}{N} + \frac{N-2}{N-1} = \frac{N-2}{N} \left(\frac{1}{N-1} + \frac{1}{N-2} \right). \end{aligned}$$

Нетрудно видеть, что

$$\mathbb{E}(\eta_i) = \frac{i-1}{N} \left(\frac{1}{N-1} + \frac{1}{N-2} + \dots + \frac{1}{N-i} \right).$$

Таким образом, эта величина перестанет расти, когда $\frac{i-1}{N} \left(\frac{1}{N-1} + \frac{1}{N-2} + \dots + \frac{1}{N-i} \right)$ зашкалит за i/N . Этот момент и является оптимальным моментом, после которого может быть произведен выбор по сформулированному выше алгоритму.

§37. Стационарные случайные процессы. Прогноз

1. Определение стационарного случайного процесса.

Определение. Случайный процесс η_t называется *стационарным* (в узком смысле), если его совместные распределения не меняются со временем, то есть

$$\mathbb{P}_{t_1, \dots, t_n} = \mathbb{P}_{t_1+h, \dots, t_n+h}$$

для любых h, t_1, \dots, t_n . Заметим, что такой процесс должен быть задан не на $[0, \infty)$, а на $(-\infty, \infty)$.

Определение. Случайный процесс η_t называется *стационарным в широком смысле*, если его среднее и центральные смешанные моменты (ковариация) не меняются со временем, то есть

$$\mathbb{E}\xi_t = \mathbb{E}\xi_0 \text{ для любого } t,$$

$$\mathbb{E}[(\xi_{t+h} - \mathbb{E}\xi_{t+h})(\xi_t - \mathbb{E}\xi_t)] = K_\xi(h) \text{ для всех } t \text{ и } h. \quad (1)$$

В мировой литературе *нормальное* распределение часто называется *гауссовским*, а случайный процесс с совместными гауссовскими распределениями — *гауссовским случайнм процессом*. Винеровский процесс является гауссовским. Мы знаем, что если все совместные распределения процесса нормальны (то есть гауссовые), то, для их описания достаточно знать векторы средних и матрицы вторых моментов. Поэтому стационарный в широком смысле гауссовский процесс является стационарным и в узком смысле.

Из стандартного винеровского процесса $w(t)$ легко получить стационарный процесс, который носит имя процесса Орнштейна – Уленбека:

$$\xi_t := e^t w(e^{-2t}).$$

Очевидно, что средние процесса Орнштейна – Уленбека равны нулю. Проверим инвариантность ковариации процесса при сдвиге времени:

$$\mathbb{E}(\xi_t \xi_{t+h}) = e^t e^{t+h} w(e^{-2t}) w(e^{-2t-2h}) = e^t e^{t+h} \min\{e^{-2t}, e^{-2t-2h}\} = e^{-h}, \text{ если } h > 0.$$

Прежде чем перейти к другим примерам гауссовых стационарных процессов, рассмотрим случайное колебание

$$\xi_t = Z \cos \lambda t, \text{ где } Z \text{ — центрированная гауссовская случайная величина, } \lambda \in \mathbf{R}.$$

Очевидно, что такой процесс не удовлетворяет условию (1), более того:

$$\mathbb{E}[(\xi_t - \mathbb{E}\xi_t)(\xi_t - \mathbb{E}\xi_t)] = \cos^2(\lambda t) \mathbb{E}Z^2 (\text{ зависит от } t).$$

Но если перенести определение стационарности на комплексные случайные процессы, то процесс ξ_t можно представить как вещественную часть стационарного процесса ζ_t :

$$\zeta_t = Z e^{i\lambda t}.$$

При этом соотношение (1) мы представляем как частный случай соотношения

$$\mathbb{E}[(\xi_{t+h} - \mathbb{E}\xi_{t+h})(\overline{\xi_t - \mathbb{E}\xi_t})] = K_\xi(h) \text{ для всех } t \text{ и } h. \quad (1')$$

Действительно,

$$K_\zeta(h) = \mathbb{E}\zeta_t \overline{\zeta_{t+h}} = e^{i\lambda t} e^{-i\lambda(t+h)} \mathbb{E}Z^2 = e^{-i\lambda h} \mathbb{E}Z^2.$$

Эту конструкцию можно обобщить, рассмотрев сумму случайных колебаний с разными частотами λ_i :

$$\zeta_t = \sum_k Z_k e^{i\lambda_k t},$$

где Z_k — центрированные независимые гауссовые случайные величины. (Независимость для стационарности не нужна, но оказывается, что подобное представление имеет место для любого стационарного гауссовского процесса ξ_t :

$$\xi_t = \int_{\mathbf{R}} e^{i\lambda t} dZ(\lambda), \quad (*)$$

где Z — случайная мера на борелевской σ -алгебре \mathbf{R} , значения которой — центрированные гауссовские случайные величины, причем для непересекающихся множеств A и B случайные величины $Z(A)$ и $Z(B)$ независимы и (это мера) $Z(A + B) = Z(A) + Z(B)$. Доказательство представления (*) трудоемко и опирается на классическую теорему Боннера.

2. Теорема Боннера.

Прежде чем формулировать эту теорему, мы введем важнейшее свойство неотрицательной определенности функции и установим справедливость этого свойства для ковариации K_ξ .

Определение. Комплексная функция K на \mathbf{R} называется *неотрицательно определенной*, если

$$\sum_{j,k} K(t_j - t_k) \sigma_j \overline{\sigma_k} \geq 0$$

для любых конечных наборов одинаковой мощности $\{t_j\} \subset \mathbf{R}$ и $\{\sigma_j\} \subset \mathbf{C}$. Аналогично определяется понятие *неотрицательной определенности* функции на множестве целых чисел \mathbf{Z} , но в этом случае $\{t_j\} \subset \mathbf{Z}$.

Теорема А. i) Пусть $K : \mathbf{R} \rightarrow \mathbf{C}$ — комплекснозначная непрерывная функция, имеющая свойство неотрицательной определенности. Тогда K представима в виде

$$K(t) = \int_{\mathbf{R}} e^{i\lambda t} d\mu(\lambda),$$

где μ — конечная мера на борелевской σ -алгебре \mathbf{R} , причем $\mu(\mathbf{R}) = K(0)$.

ii) Пусть $K : \mathbf{Z} \rightarrow \mathbf{C}$ — неотрицательно определенная функция. Тогда K представима в виде

$$K(n) = \int_{[-\pi, \pi)} e^{i\lambda n} d\mu(\lambda),$$

где μ — конечная мера на борелевской σ -алгебре $[-\pi, \pi)$, причем $\mu([-\pi, \pi)) = K(0)$.

Частным случаем этой теоремы является теорема Боннера для характеристических функций. Впрочем, теорема А сводится к теореме В делением на $K(0)$. Нам будет удобнее доказать теорему В, так как при ее доказательстве мы сможем использовать известные нам факты о слабой сходимости распределений.

Теорема В. Функция $\varphi : \mathbf{R} \rightarrow \mathbf{C}$ является характеристической функцией распределения \mathbb{P} (m.e.)

$$\varphi(t) = \int_{\mathbf{R}} e^{itx} d\mathbb{P}(x)) \tag{1}$$

тогда и только тогда, когда

- i) $\varphi(0) = 1$,
- ii) φ неотрицательно определена,
- iii) φ непрерывна в точке 0.

Прежде чем доказывать теорему В, проверим свойство неотрицательной определенности для описываемых в этих теоремах объектов.

Предложение А. Ковариация K_ξ стационарного случайного процесса ξ_t неотрицательно определена.

Доказательство. Для упрощения записи будем считать, что среднее процесса равно нулю (его вычли). Положим в (1') $t = t_k$, $t + h = t_j$. Имеем:

$$\sum_{j,k} K(t_j - t_k) \sigma_j \overline{\sigma_k} = \sum_{j,k} \mathbb{E} (\xi_{t_j} \overline{\xi_{t_k}} \sigma_j \overline{\sigma_k}) = \mathbb{E} \left| \sum_j \xi_{t_j} \sigma_j \right|^2 \geq 0.$$

Предложение B. Характеристическая функция φ_ξ случайной величины ξ неотрицательно определена.

Доказательство.

$$\sum_{j,k} \varphi_\xi(t_j - t_k) \sigma_j \bar{\sigma_k} = \sum_{j,k} \mathbb{E} \left(e^{i\xi t_j} e^{-i\xi t_k} \sigma_j \bar{\sigma_k} \right) = \mathbb{E} \left| \sum_j e^{i\xi t_j} \sigma_j \right|^2 \geq 0.$$

Доказательство теоремы B. 'Только тогда'. Свойства i) и iii) характеристической функции см. в курсе теории вероятностей. Свойство ii) доказано в предложении B.

Доказательство 'тогда' значительно сложнее и использует слабую компактность пространства всех вероятностных мер на компакте K , которая следует из теоремы Рисса для пространства $C(K)$ — пространство вероятностных мер замкнуто в слабой топологии и находится внутри единичного шара сопряженного к $C(K)$ пространства зарядов, который по теореме Алаоглу слабо компактен. Для случая, когда K — отрезок вещественной прямой, теорема Алаоглу не нужна, слабую компактность пространства вероятностных мер мы доказали.

Лемма 1. Для всех t и s из \mathbf{R}

- a) $\varphi(-t) = \bar{\varphi}(t)$,
- b) $|\varphi(t)| \leq \bar{\varphi}(0) = 1$,
- c) $|\varphi(t) - \varphi(s)|^2 \leq 1 - |\varphi(t-s)|^2 - 2\Re[\varphi(t)\bar{\varphi}(s)(1 - \varphi(t-s))]$.

Из неравенства c) и свойства iii) следует равномерная непрерывность характеристической функции (проверить!).

Доказательство сводится к подбору соответствующих вариантов наборов в (2).

Для доказательства a) надо взять $\{t_1, t_2\} = \{0, t\}$, а для $\{\sigma_1, \sigma_2\}$ использовать два варианта: $\{1, 1\}$ и $\{1, i\}$. Первый вариант позволяет получить a) для мнимых частей $\varphi(t)$ и $\varphi(-t)$, а второй вариант — для вещественных частей этих чисел. (Проверить!)

Для доказательства b) надо взять также $\{t_1, t_2\} = \{0, t\}$, но далее рассматривать (2) как квадратичную форму для переменных $\{\sigma_1, \sigma_2\}$ и использовать неотрицательность определителя этой формы. (Проверить!)

Для доказательства c) мы берем наборы $\{t_1, t_2, t_3\} = \{0, t, s\}$ и опять используем свойство квадратичной формы (2) от переменных $\{\sigma_1, \sigma_2, \sigma_3\}$ — неотрицательность определителя третьего порядка ((проверить!))

$$\begin{vmatrix} 1 & \varphi(t) & \varphi(s) \\ \bar{\varphi}(t) & 1 & \varphi(t-s) \\ \bar{\varphi}(s) & \bar{\varphi}(t-s) & 1 \end{vmatrix} \geq 0.$$

Лемма 2 (Герглотц). Функция ψ неотрицательна определена на группе по сложению

$$D = \{\dots - 3c, -2c, -c, 0, c, 2c, 3c \dots\}, c > 0,$$

тогда и только тогда, когда она совпадает на G с характеристической функцией вероятностной меры на множестве $[-\pi/c, +\pi/c]$:

$$\psi(mc) = \int_{[-\pi/c, +\pi/c]} e^{imcx} dF(x). \quad (3)$$

Доказательство. Сужение функции ψ на конечное множество $[-nc, nc] \cap D$ легко представляется в виде (3): мы вводим плотность

$$G'_n(x) := \frac{1}{2\pi} \sum_{m=-n+1}^{n-1} \left(1 - \frac{|m|}{n} \right) \psi(mc) e^{-imx} = (?)$$

$$= \frac{1}{2\pi n} \sum_{j=1}^n \sum_{k=1}^n \psi((j-k)c) e^{-i(j-k)x} \geq 0.$$

Умножая это равенство на e^{imx} при фиксированном m и интегрируя по $[-\pi, +\pi]$, мы получаем (?)

$$\left(1 - \frac{|m|}{n}\right) \psi(mc) = \int_{[-\pi, +\pi)} e^{imx} G'_n(x) dx = \int_{[-\pi/c, +\pi/c)} e^{imcx} dF_n(x), \quad (4)$$

где F_n — функция распределения, причем $F_n(-\pi/c) = 0$, $F_n(\pi/c) = 1$. (Проверить!)

Далее мы используем слабую компактность последовательности мер dF_n и извлекаем из нее сходящуюся подпоследовательность $dF_{n_k} \xrightarrow{\text{сл}} dF$. Переходя к пределу по k в (4), получаем (1). (Проверить!)

Теперь перейдем к доказательству теоремы Бонхера. Мы будем доказывать, что φ совпадает с характеристической функцией некоторого распределения на всюду плотном множестве $\tilde{\mathbf{Q}}$ всех двоично рациональных чисел в \mathbf{R} . Обозначим через Q_n множество всех рациональных чисел вида $k/2^n$, где k — целое число, n — натуральное число. Тогда $Q_n \uparrow \tilde{\mathbf{Q}}$. Очевидно, что из неотрицательной определенности φ на группе по сложению \mathbf{R} следует неотрицательная определенность сужения φ на каждую подгруппу Q_n . Каждому такому сужению применима лемма 2 и по ней для каждого n на множестве $[-2^n\pi, 2^n\pi)$ существует такая вероятностная мера dF_n , что

$$\varphi(k/2^n) = \int_{[-2^n\pi, 2^n\pi)} e^{i(k/2^n)x} dF_n(x) \text{ для всех целых } k. \quad (5)$$

Введем характеристические функции ψ_n распределений dF_n . Из (5) следует, что $\psi_n \rightarrow \varphi$ поточечно на множестве $\tilde{\mathbf{Q}}$. Далее нашей целью будет доказательство равностепенной непрерывности последовательности характеристических функций (ψ_n) в точке 0. Из нее (точно так же, как это делалось в доказательстве теоремы непрерывности (посмотреть!)), будет следовать слабая компактность последовательности мер (dF_n).

Для характеристических функций ψ вероятностных распределений на \mathbf{R} справедливо trivialно доказываемое с помощью неравенства Шварца

Неравенство для приращений.

$$|\psi(t) - \psi(t+h)|^2 \leq 2|1 - \Re\psi(h)|.$$

Доказательство неравенства. Напомним, что $\psi(t) = \int_{\mathbf{R}} e^{itx} d\mathbb{P}(x)$, где \mathbb{P} — распределение вероятностей на \mathbf{R} . Имеем:

$$\begin{aligned} |\psi(t) - \psi(t+h)|^2 &= \left\{ \int_{\mathbf{R}} (e^{itx} - e^{i(t+h)x}) d\mathbb{P}(x) \right\}^2 = \left\{ \int_{\mathbf{R}} (e^{itx} [1 - e^{ihx}]) d\mathbb{P}(x) \right\}^2 \leq \\ &\leq \int_{\mathbf{R}} |e^{itx}|^2 d\mathbb{P}(x) \cdot \int_{\mathbf{R}} |1 - e^{ihx}|^2 d\mathbb{P}(x) = 1 \cdot \int_{\mathbf{R}} [1 - e^{ihx} - e^{-ihx} + 1] d\mathbb{P}(x) = \\ &= \int_{\mathbf{R}} [2 - \cos(hx)] d\mathbb{P}(x) = 2(1 - \Re\psi(h)). \end{aligned}$$

Вернемся к доказательству теоремы Бонхера. Возьмем $\theta \in [0, 1/2^n]$. Следующее неравенство основано на монотонности \cos на $[0, \pm\pi]$ и является решающим:

$$1 - \Re\psi_n(\theta/2^n) = \int_{[-\pi, \pi)} (1 - \cos(\theta x)) dF_n(2^n x) \leq$$

$$\leq \int_{[-\pi, \pi]} (1 - \cos x dF_n(2^n x) = 1 - \Re \varphi(1/2^n).$$

Далее каждое h в окрестности нуля для каждого n мы представляем в виде $h = (k_n + \theta_n)/2^n$ ($0 \leq \theta_n < 1$, k_n — целое) и оцениваем расстояние $|1 - \psi_n(h)|^2$ с помощью неравенства треугольника:

$$\begin{aligned} |1 - \psi_n(h)|^2 &\leq 2|1 - \psi_n(k_n/2^n)|^2 + 2|\psi_n(k_n/2^n) - \psi_n(h)|^2 \leq \\ &\leq 2|1 - \psi_n(k_n/2^n)|^2 + 4(1 - \Re \psi_n(\theta_n/2^n))^2 \leq 2|1 - \psi_n(k_n/2^n)|^2 + 4(1 - \Re \psi_n(1/2^n))^2 = \\ &= 2|1 - \varphi(k_n/2^n)|^2 + 4(1 - \Re \varphi(1/2^n))^2. \end{aligned}$$

Используя непрерывность φ , правую часть можно сделать сколь угодно малой за счет выбора окрестности для h . Итак, из непрерывности φ в нуле следует равностепенная непрерывность последовательности (ψ_n) в нуле. Далее мы выбираем слабо сходящуюся подпоследовательность функций распределения F_{n_k} , предел которой F пока не обязан быть функцией распределения. Доказательство того, что это — функция распределения, проводится так же, как и в доказательстве теоремы непрерывности. Небольшое отличие состоит в том, что неравенство для усечений используется здесь для каждой характеристической функции ψ_n , при этом нужная оценка получается ввиду равностепенной непрерывности.

Замечание к теореме непрерывности. Легко проверяется, что в формулировке теоремы непрерывности условие того, что предельная функция φ является характеристической функцией, излишне. Достаточно потребовать непрерывность функции φ в точке 0. А именно, сначала доказывается слабая компактность семейства функций распределений F_n , в качестве F берется предел некоторой сходящейся подпоследовательности F_{n_k} , после чего доказывается, что $\varphi = \varphi_F$, $F_n \xrightarrow{w} F$.

Завершим это доказательства формулировкой теоремы, к которой мы стремились:

Теорема А. i) Пусть ξ_t — стационарный непрерывный гауссовский процесс на \mathbf{R} (или \mathbf{Z}), $K_\xi : \mathbf{R} \rightarrow \mathbf{C}$ (или $K_\xi : \mathbf{R} \rightarrow \mathbf{C}$) ковариация процесса ξ . Тогда имеет место представление

$$K(t) = \int_{\mathbf{R}} e^{i\lambda t} d\mu_\xi(\lambda),$$

где μ_ξ — конечная мера на борелевской σ -алгебре \mathbf{R} , причем $\mu(\mathbf{R}) = K(0)$ (или

$$K(n) = \int_{[-\pi, \pi]} e^{i\lambda n} d\mu_\xi(\lambda),$$

где μ_ξ — конечная мера на борелевской σ -алгебре $[-\pi, \pi]$, причем $\mu_\xi([-\pi, \pi]) = K(0)$.

Мы будем называть μ_ξ спектральной мерой процесса ξ , а если

$$d\mu_\xi(\lambda) = f_\xi(\lambda) d\lambda,$$

то функцию f_ξ будем называть спектральной плотностью процесса.

4. Спектральное разложение для стационарных случайных процессов.

Мы будем в дальнейшем рассматривать лишь гауссовские стационарные процессы, причем непрерывные в среднем квадратическом.

Теорема. Произвольный гауссовский стационарный случайный процесс ξ_t ($t \in \mathbf{R}$) можно представить в виде:

$$\xi_t = \int_{\mathbf{R}} e^{i\lambda t} dZ(\lambda), \tag{*}$$

где Z — случайная мера на борелевской σ -алгебре \mathbf{R} , значения которой — центрированные гауссовые случайные величины, причем для непересекающихся множеств A и B случайные величины $Z(A)$ и $Z(B)$ независимы и (это — мера) $Z(A+B) = Z(A) + Z(B)$.

Доказательство представления (*) опирается на теорему Боннера. Здесь мы объясним, как строится случайная мера Z . Отображение Φ , переводящее случайные величины ξ_t в функции $\lambda \rightarrow e^{i\lambda t} \in L^2(\mathbf{R}, \mathfrak{B}(\mathbf{R}), \mu)$, является изометрическим и продолжается до изометрического отображения на замыкание $\text{Lin}\{\xi_t : t \in \mathbf{R}\}$ в пространстве $L^2((\Omega, \mathfrak{A}, \mathbb{P})$. Изометричность следует из определения спектральной меры μ . Легко показывается, что $\text{Lin}\{\lambda \rightarrow e^{i\lambda t} : t \in \mathbf{R}\}$ всюду плотно в $L^2(\mathbf{R}, \mathfrak{B}(\mathbf{R}), \mu)$. Мы берем $Z(B) := \Phi^{-1}(I_B)$.

Для обоснования этой конструкции нужно проверить, что функции $\lambda \rightarrow e^{i\lambda t}$ порождают $L^2(\mathbf{R}, \mathfrak{B}(\mathbf{R}), \mu)$. Для этого достаточно проверить, что линейные комбинации функций $e^{i\lambda t}$ в метрике $L^2(\mathbf{R}, \mathfrak{B}(\mathbf{R}), \mu)$ индикаторы всех борелевских множеств. Более того, в силу счетной аддитивности и ограниченности μ нам достаточно показать возможность приближения индикаторов $I_{(-\infty, a]}$ на любом отрезке $[b, c]$. Из теории рядов Фурье или теоремы Стоуна—Вейерштрасса следует возможность равномерного приближения любой непрерывной функции. В свою очередь, как мы это уже делали (где?) ограниченная последовательность непрерывных функций поточечно сходится к $I_{(-\infty, a]}$.

5. Прогноз для процессов с дискретным временем

Для упрощения ситуации мы будем рассматривать далее стационарные процессы с дискретным временем $\{n \in bfN\}$. В этом случае спектральная мера процесса сосредоточена на окружности $\lambda \in [-\pi, \pi]$ и имеется по крайней мере один стационарный процесс, для которого спектральная мера вычисляется. Этот процесс η_n мы будем называть *процессом белого шума*, его ковариация $K(n)$ равна $I_{\{0\}}(n)$. Таким образом случайные величины η_n независимы, имеют среднее 0 и дисперсию 1. Очевидно, что

$$K(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} 1 \cdot e^{-n\lambda} d\lambda,$$

таким образом спектральная мера μ_η имеет плотность $\frac{1}{2\pi}$. Далее предлагается использовать модели, использующие стационарные процессы, получающиеся из белого шума некоторыми линейными преобразованиями и для них строить теорию оптимального прогноза (подробное изложение см. в [4]).

Давайте рассмотрим два стационарных процесса, ξ_n и ζ_n , связанные линейными соотношением:

$$\zeta_n = a_0 \xi_n + a_1 \xi_{n-1} + \dots + a_m \xi_{n-m},$$

оказывается, что по спектральной плотности одного из этих процессов можно восстановить спектральную плотность другого процесса. Дело в том, что если случайные величины ξ_n мы отождествляем с функциями $e^{in\lambda}$, то тогда случайные величины ζ_n следуют отождествлять с функциями $a_0 e^{in\lambda} + a_1 e^{i(n-1)\lambda} + \dots + a_m e^{i(n-m)\lambda}$. Теперь подсчитаем K_ζ :

$$\begin{aligned} K_\zeta(n) &= \mathbb{E}[\zeta_0 \overline{\zeta_n}] = \\ &= \int_{-\pi}^{\pi} (a_0 e^{i0\lambda} + a_1 e^{-i\lambda} + \dots + a_m e^{-im\lambda})(\overline{a_0 e^{in\lambda} + a_1 e^{i(n-1)\lambda} + \dots + a_m e^{i(n-m)\lambda}}) f_\xi(\lambda) d\lambda = \\ &= \int_{-\pi}^{\pi} e^{i0\lambda} \overline{e^{in\lambda}} (a_0 e^{i0\lambda} + a_1 e^{-i\lambda} + \dots + a_m e^{-im\lambda})(\overline{a_0 e^{i0\lambda} + a_1 e^{-i\lambda} + \dots + a_m e^{-im\lambda}}) f_\xi(\lambda) d\lambda = \\ &= \int_{-\pi}^{\pi} e^{i0\lambda} \overline{e^{in\lambda}} f_\xi(\lambda) d\lambda. \end{aligned}$$

Но тогда мы можем положить

$$f_\zeta(\lambda) = f_\xi(\lambda) |a_0(e^{i0\lambda} + a_1 e^{i\lambda} + \dots + a_m e^{im\lambda})|^2.$$

Теперь мы можем выписать спектральную плотность для двух видов процессов, которые определяются с помощью процесса белого шума: *процесс скользящего среднего* ξ_n определяется соотношением

$$\xi_n = a_0 \eta_n + a_1 \eta_{n-1} + \dots + a_m \eta_{n-m},$$

процесс авторегрессии ζ_n определяется обратным соотношением

$$\eta_n = b_0\zeta_n + b_1\zeta_{n-1} + \dots + b_m\zeta_{n-k}. \quad (*)$$

Согласно только сформулированному принципу получаем:

$$f_\xi(\lambda) = \frac{1}{2\pi} |a_0(e^{i0\lambda} + a_1e^{i\lambda} + \dots + a_m e^{im\lambda})|^2, \quad f_\zeta(\lambda) = \frac{1}{2\pi} \frac{1}{|b_0(e^{i0\lambda} + b_1e^{i\lambda} + \dots + b_k e^{ik\lambda})|^2}.$$

Отметим, что из тех же соображений мы можем выписать спектральную плотность процесса, которая получается из процесса белого шума в два этапа, сначала по процессу η_n определяется процесс ζ_n , потом по нему строится соотношением (*) процесс ζ_n , но в (*) роль η играет процесс ξ . В этой конструкции спектральная плотность является дробью

$$\frac{1}{2\pi} \frac{|a_0(e^{i0\lambda} + a_1e^{i\lambda} + \dots + a_m e^{im\lambda})|^2}{|b_0(e^{i0\lambda} + b_1e^{i\lambda} + \dots + b_k e^{ik\lambda})|^2}. \quad (1)$$

Теперь, наконец, перейдем к задаче прогноза.

Решением задачи прогноза для стационарного процесса ξ_n на один шаг является случайная величина $\hat{\xi}_1 \in \overline{\text{Lin}\{\xi_0, \xi_{-1}, \dots\}}$ (замыкание в норме L^2), которая минимизирует $\mathbb{E}(\xi_1 - \hat{\xi}_1)^2$. Хорошо известно хотя бы из трехмерного опыта, что такой минимум достигается при ортогональном проектировании ξ_1 на $\overline{\text{Lin}\{\xi_0, \xi_{-1}, \dots\}}$, а это в свою очередь эквивалентно ортогональности $\xi_1 - \hat{\xi}_1$ пространству $\overline{\text{Lin}\{\xi_0, \xi_{-1}, \dots\}}$, а для этого достаточно $\xi_1 - \hat{\xi}_1 \perp \xi_n$ для любого $n \in \{0, -1, -2, \dots\}$.

Теперь перенесем эту ситуацию в пространство $H_\xi = L^2([-\pi, \pi], f_\xi(\lambda)d\lambda)$, где f_ξ — спектральная плотность процесса ξ . При таком переходе случайные величины ξ_n , где $n \in \mathbf{Z}$, \mathbf{Z} — множество всех целых чисел, переходят в функции $\lambda \rightarrow e^{in\lambda}$, пространство $\overline{\text{Lin}\{\xi_0, \xi_{-1}, \dots\}}$ переходит в подпространство $H_\xi = L^2([-\pi, \pi], f_\xi(\lambda)d\lambda)$, образ случайной величины $\hat{\xi}_1$ мы обозначим

$$g \in \overline{\text{Lin}\{1 = e^{i0\lambda}, e^{-i\lambda}e^{-i2\lambda}, \dots\}},$$

причем должно выполняться $e^{i\lambda} - g \perp e^{in\lambda}$ в $H_\xi \forall n = 0, -1, -2, \dots$, то есть

$$\int_{-\pi}^{\pi} (e^{i\lambda} - g(\lambda)) \overline{e^{in\lambda}} f_\xi(\lambda) d\lambda = 0 \quad \forall n = 0, -1, -2, \dots \quad (*)$$

Далее мы покажем, как решается задача прогноза лишь для конкретного примера $f_\xi(\lambda) = (2 + e^{-i\lambda})(2 + e^{i\lambda})$. Заметим, что имеет место такая *факторизация* $f_\xi = f_1 f_2$ (где $f_1 = 2 + e^{-i\lambda}$, $f_2 = 2 + e^{i\lambda}$), что функции f_1 и f_2^{-1} являются суммами равномерно сходящегося ряда по функциям $e^{in\lambda}, n = 0, -1, -2, \dots$ (причем для f_1 это просто сумма), аналогично функции f_2 и f_2^{-1} являются суммами равномерно сходящегося ряда по функциям $e^{in\lambda}, n = 0, 1, 2, \dots$. Следующую лемму мы предлагаем доказать самостоятельно:

Лемма. Пусть функция u принадлежит $\overline{\text{Lin}\{e^{-i\lambda}, e^{-i2\lambda}, \dots\}}$, то есть

$$u(\lambda) = \sum_{n>0} a_n e^{in\lambda}, \quad \text{сходимость в пространстве } L^2([-\pi, \pi], d\lambda), \quad (**)$$

функция v разлагается в равномерно сходящийся ряд

$$v(\lambda) = \sum_{n \geq 0} b_n e^{in\lambda}, \quad \text{сходимость в пространстве } C([-\pi, \pi]),$$

тогда uv разлагается в ряд

$$u(\lambda)v(\lambda) = \sum_{n>0} c_n e^{in\lambda}, \quad \text{сходимость в пространстве } L^2([-\pi, \pi], d\lambda).$$

Здесь $d\lambda$ — обычная мера Лебега без умножения на спектральный множитель.

Замечание. Условие равномерной сходимости v по-видимому существенно. В противном случае просто может не получиться принадлежности $L^2([-\pi, \pi], d\lambda)$. К сожалению, не могу предложить контрпример.

Итак, как использовать определенное выше разложение. Для наглядности введем четыре замкнутых линейных подпространства пространства $L^2([-\pi, \pi], d\lambda)$:

$$\begin{aligned} H_+ &:= \left\{ u|[-\pi, \pi] : u = \sum_{n \geq 0} a_n e^{in\lambda}, \sum a_n^2 < \infty \right\}, \\ \tilde{H}_+ &:= \left\{ u|[-\pi, \pi] : u = \sum_{n > 0} a_n e^{in\lambda}, \sum a_n^2 < \infty \right\}, \\ H_- &:= \left\{ u|[-\pi, \pi] : u = \sum_{n \geq 0} a_n e^{-in\lambda}, \sum a_n^2 < \infty \right\}, \\ \tilde{H}_- &:= \left\{ u|[-\pi, \pi] : u = \sum_{n > 0} a_n e^{-in\lambda}, \sum a_n^2 < \infty \right\}. \end{aligned}$$

Очевидно, что

$$H_+ \perp \tilde{H}_- H_- \perp \tilde{H}_+, L^2([-\pi, \pi], d\lambda) = H_+ \oplus \tilde{H}_- = H_- \oplus \tilde{H}_+,$$

\oplus здесь означает ортогональную прямую сумму. Соотношение (*) мы можем представить как

$$(e^{i\lambda} - g(\lambda))f_1(\lambda)f_2(\lambda) \perp e^{in\lambda} \forall n = 0, -1, -2, \quad (***)$$

или (что то же) $(e^{i\lambda} - g(\lambda))f_1(\lambda)f_2(\lambda) \in \tilde{H}_+$. Согласно лемме

$$(e^{i\lambda} - g(\lambda))f_1(\lambda) = (e^{i\lambda} - g(\lambda))f_1(\lambda)f_2(\lambda)f_2^{(-1)}(\lambda) \in \tilde{H}_+.$$

Обратно, из этого соотношения следует (***)�. Дальнейшее просто: мы берем ряд $e^{i\lambda}f_1(\lambda)$ и берем из него ту часть ряда, которая относится к H_- , она равна $g(\lambda)f_1(\lambda)$, для вычисления g нужно умножить $g(\lambda)f_1(\lambda)$ на $f_1^{-1}(\lambda)$.

Итак,

$$\begin{aligned} e^{i\lambda}f_1(\lambda) &= 2e^{i\lambda} + 1, \\ g(\lambda)f_1(\lambda) &= 1, \\ g(\lambda) &= 1/f_1(\lambda) = 1/2e^{0i\lambda} - 1/4e^{-i\lambda} + 1/8e^{-2i\lambda} - 1/16e^{-3i\lambda} + \dots. \end{aligned} \quad (****)$$

Переходя к случайному процессу, получаем

$$\hat{\xi}_1 = \xi_0/2 - \xi_{-1}/4 + \xi_{-2}/8 - \xi_{-3}/16 + \dots$$

То, что в правой части (****) мы получили 1, объясняется простотой процесса. Впрочем, если бы мы прогнозировали на два шага, мы просто получили бы 0, то есть прогноз на два шага ничего не дает кроме среднего процесса.

Наконец, что делать для более сложных процессов, как получить факторизацию $f = f_1f_2$ даже для процесса со спектральной плотностью $(1 + 2e^{-i\lambda})(1 + 2e^{i\lambda})$. В этом случае деление 1 на каждый из сомножителей не приводит к равномерно сходящемуся ряду по степеням $e^{in\lambda}$. Оказывается, все просто, умножив $1 + 2e^{-i\lambda}$ на $e^{i\lambda}$, а $1 + 2e^{i\lambda}$ на $e^{-i\lambda}$, мы приходим к старой ситуации. Аналогично действуя и используя основную теорему алгебры, мы можем доказать существование факторизации для любой функции вида (1).

3. Теорема. Имеется стационарный случайный процесс ξ_n , $n \in \mathbf{Z}$, с нулевым средним и с дискретным временем. Тогда на единичной окружности существует такая случайная мера Z с ортогональными приращениями, что

$$\xi_n = \int_{[-\pi, \pi)} e^{i\lambda n} dZ(\lambda). \quad (1)$$

Из стационарности процесса следует неотрицательная определенность функции

$$K(n) = \mathbf{E}\xi_0\bar{\xi}_n.$$

Действительно

$$\begin{aligned} \sum_m \sum_n K(m-n)\sigma_m\bar{\sigma}_n &= \mathbf{E} \left(\sum_m \sum_n \xi_m \bar{\xi}_n \sigma_m \bar{\sigma}_n \right) = \\ &= \mathbf{E} \left| \sum_m \xi_m \sigma_m \right|^2 \geq 0. \end{aligned}$$

Заметим, что по любой возрастающей функции F на $[\pi, \pi)$ можно построить случайную меру Z с ортогональными приращениями с гауссовскими распределениями, такую, что $F(\lambda) = EZ^2(\lambda)$. Это следует из теоремы Колмогорова о совместных распределениях.

Введем пространство H^2 — замыкание $\text{Lin}\{\xi_n : n \in \mathbf{Z}\}$ в метрике, заданной скалярным произведением

$$(\xi, \eta) := \mathbf{E}(\xi\bar{\eta}).$$

Одновременно рассмотрим функцию $F(\lambda)$, существование которой следует из теоремы Боннера и которая определяется соотношением

$$K(n) = \int_{[-\pi, \pi)} e^{i\lambda n} dF(\lambda), \mathbf{E}(\xi_m \bar{\xi}_n) = K(m-n) = \int_{[-\pi, \pi)} e^{i\lambda(m-n)} dF(\lambda).$$

Рассмотрим отображение $\Phi : \xi_n \rightarrow e^{i\lambda n}$, которое является изометрическим отображением H^2 в пространство $L^2([-\pi, \pi), dF)$. Легко видеть (??1), что это отображение изометрическим, а согласно теореме Стоуна — Вейерштрасса образ этого отображения всюду плотен. Обратное отображение переводит функции $I_{[0, \lambda]}$ в некоторые случайные величины $Z(\lambda)$. Ортогональность приращений $Z(\mu) - Z(\lambda)$, $Z(\lambda)$, $0 < \lambda < \mu$ следует из ортогональности функций $I_{[0, \lambda]}$ и $I_{[\lambda, \mu]}$ в пространстве $L^2([-\pi, \pi), dF)$.

Функция F может являться интегралом от некоторой функции f , которая называется *спектральной плотностью*. Как находить спектральную плотность для некоторых случайных процессов, задаваемых некоторыми условиями, не используя обратное преобразование Фурье? Надо начинать от простейшего стационарного процесса ζ_n , состоящего из независимых $N(0, 1)$ -нормальных случайных величин. В этой ситуации

$$K_\zeta(n) = \begin{cases} 1, & \text{если } n = 0, \\ 0, & \text{если } n \neq 0. \end{cases}$$

Легко видеть, что спектральная плотность ζ_n имеет вид $f(\lambda) = 1/2\pi$, так как из теории рядов Фурье известно, что функции $e^{i\lambda n}$ ортогональны для разных n относительно меры $d\lambda$. Так как все точки спектра с равной интенсивностью входят в спектр, процесс ζ_n называется *процессом белого шума*. Для этого процесса задача прогноза не имеет смысла, так как будущее не зависит от прошлого и настоящего, следовательно, и прогноз не будет зависеть от прошлого и настоящего.

Теперь посмотрим, что происходит со спектральной плотностью f_φ процесса при линейном преобразовании произвольного процесса φ_n . А именно, от процесса φ_n мы переходим к процессу $\psi_n = a_0\varphi_n + a_1\varphi_{n-1} + \dots + a_k\varphi_{n-k}$. Однако при построенном ранее отображении случайные

величины ψ_n переходят в функции $a_0e^{i\lambda n} + a_1e^{i\lambda(n-1)} + \dots + a_ke^{i\lambda(n-k)}$, в частности, ψ_0 переходит в функцию $a_0e^{i\lambda 0} + a_1e^{i\lambda(-1)} + \dots + a_ke^{i\lambda(-k)}$, тогда корреляционная функция процесса ψ_n также интерпретируется как скалярное произведение этих функций и будет иметь вид

$$\begin{aligned} K_\psi(n) &= \int_{[-\pi, \pi)} e^{i\lambda n} f_\psi(\lambda) d\lambda = \\ &= \int_{[-\pi, \pi)} (a_0e^{i\lambda n} + a_1e^{i\lambda(n-1)} + \dots + a_ke^{i\lambda(n-k)}) \overline{(a_0e^{i\lambda 0} + a_1e^{i\lambda(-1)} + \dots + a_ke^{i\lambda(-k)})} f_\varphi(\lambda) d\lambda = \\ &= \int_{[-\pi, \pi)} e^{i\lambda n} \left| a_0e^{i\lambda 0} + a_1e^{i\lambda(-1)} + \dots + a_ke^{i\lambda(-k)} \right|^2 f_\xi(\lambda) d\lambda. \end{aligned}$$

Отсюда немедленно следует соотношение между f_φ и f_ψ :

$$f_\psi(\lambda) = \left| a_0e^{i\lambda 0} + a_1e^{i\lambda(-1)} + \dots + a_ke^{i\lambda(-k)} \right|^2 f_\varphi(\lambda).$$

Теперь мы можем записать спектральную плотность трех типов процессов, которые линейным образом записываются через заданный процесс белого шума ζ_n , все три процесса мы будем обозначать одной и той же буквой ξ .

Процесс скользящего среднего — это $\xi_n = a_0\zeta_n + a_1\zeta_{n-1} + \dots + a_k\zeta_{n-k}$. Согласно сказанному ($\varphi_n = \zeta_n$, $\psi_n = \xi_n$), в этом случае

$$f_\xi(\lambda) = \frac{1}{2\pi} \left| a_0e^{i\lambda 0} + a_1e^{i\lambda(-1)} + \dots + a_ke^{i\lambda(-k)} \right|^2.$$

Процесс авторегрессии — это процесс ξ_n , связанный с процессом ζ_n следующим соотношением: $\xi_n = a_0\xi_n + a_1\xi_{n-1} + \dots + a_k\xi_{n-k}$. В этом случае $\varphi_n = \xi_n$, $\psi_n = \zeta_n$. Поэтому

$$f_\xi(\lambda) = \frac{1}{2\pi} \frac{1}{\left| a_0e^{i\lambda 0} + a_1e^{i\lambda(-1)} + \dots + a_ke^{i\lambda(-k)} \right|^2}. \quad (2)$$

Замечание. Не для любого набора (a_n) существует процесс ξ_n . Например, если $a_0 = 1$, $a_1 = -1$, $a_n = 0$ для $n > 1$, то складывая соотношения для ζ_n получаем $\zeta_n + \dots + \zeta_1 = \xi_n - \xi_0$, $\mathbf{D}(\xi_n - \xi_0) = n \rightarrow \infty$. Это следует также из неинтегрируемости функции $f_\xi(\lambda)$ в (2)(??2)

Процесс смешанного типа — это ξ_n такой, что $\xi_n = a_0\eta_n + a_1\eta_{n-1} + \dots + a_k\eta_{n-k}$, где $\zeta_n = b_0\eta_n + b_1\eta_{n-1} + \dots + b_s\eta_{n-s}$. Аналогично предыдущим двум случаям, в этой ситуации

$$f_\xi(\lambda) = \frac{1}{2\pi} \frac{\left| a_0e^{i\lambda 0} + a_1e^{i\lambda(-1)} + \dots + a_ke^{i\lambda(-k)} \right|^2}{\left| b_0e^{i\lambda 0} + b_1e^{i\lambda(-1)} + \dots + b_se^{i\lambda(-s)} \right|^2}. \quad (3)$$

Задача прогноза. Теперь мы рассмотрим задачу прогноза для стационарного случайного процесса. Идея прогноза (для простоты рассмотрим прогноз на один шаг) состоит в ортогональном проектировании случайной величины ξ_1 на замыкание линейной оболочки случайных величин ξ_0, ξ_{-1}, \dots . Действительно, как хорошо известно из геометрии, при ортогональном проектировании мы получаем элемент, находящийся от проектируемого на минимальном расстоянии. В результате проектирования мы получим вообще говоря бесконечную линейную комбинацию $c_0\xi_0 + c_1\xi_1 + \dots$. На практике это используется следующим образом: из полученных нами в результате наблюдения или эксперимента чисел $\xi_0(\omega), \xi_{-1}(\omega), \dots$ мы составляем линейную комбинацию $c_0\xi_0(\omega) + c_1\xi_{-1}(\omega) + \dots$. Полученное число будет нашим прогнозом.

Мы будем рассматривать задачу прогноза для простоты при условии $0 < C_1 \leq f_\xi(\lambda) \leq C_2 < \infty$. (??3 Почему это условие выполняется для функций из (3))?. (??4 Где это условие дальше используется?). Чтобы облегчить ситуацию, задача переносится на пространство

$L^2([-\pi, \pi], f_\xi(\lambda)d\lambda)$, в этом пространстве мы проектируем функцию $e^{i\lambda}$ на замыкание линейной оболочки функций $e^{i\lambda_0}, e^{i\lambda(-1)}, e^{i\lambda(-2)}, \dots$, в результате получаем функцию прогноза $g(\lambda) = c_0 \cdot 1 + c_{-1}e^{i\lambda(-1)} + c_{-2}e^{i\lambda(-2)}$ и нужные нам коэффициенты $c_{-i}, i = 0, 1, 2, \dots$. Иногда можно получить точные формулы для коэффициентов. Идея состоит в переходе из пространства $L^2([-\pi, \pi], f_\xi(\lambda)d\lambda)$ в пространство $L^2([-\pi, \pi], f_\xi(\lambda)d\lambda)$, в котором все просто, все функции $e^{i\lambda k}$ ортогональны при разных k .

Введем несколько линейных подпространств пространства $L^2([-\pi, \pi], d\lambda)$:

$\mathbf{h}_{\leq 0}$ является замыканием линейной оболочки функций $e^{i\lambda n}$, где $n \leq 0$, $\mathbf{h}_{>0}$ является замыканием линейной оболочки функций $e^{i\lambda n}$, где $n > 0$. Хотя сами функции $e^{i\lambda n}$ непрерывны, функции из пространств $\mathbf{h}_{\leq 0}$ и $\mathbf{h}_{>0}$ могут не быть непрерывными. Более того, каждый элемент $L^2([-\pi, \pi], d\lambda)$ однозначно раскладывается в виде суммы элемента из $\mathbf{h}_{\leq 0}$ и элемента из $\mathbf{h}_{>0}$. Эти подпространства ортогональны в смысле скалярного произведения пространства $L^2([-\pi, \pi], d\lambda)$ (но не в смысле пространства $L^2([-\pi, \pi], f_\xi(\lambda)d\lambda)$);

$\mathbf{C}_{\leq 0}$ (соответственно, $\mathbf{C}_{\geq 0}$) — пространства непрерывных функций, являющихся равномерными пределами конечных линейных комбинаций функций $e^{i\lambda n}$, где $n \leq 0$ (соответственно, $n \geq 0$). Разумеется, эти пространства незамкнуты в $L^2([-\pi, \pi], d\lambda)$. Более того, эти пространства являются не только линейными пространствами, но и алгебрами. Действительно, произведения функций $e^{i\lambda n}$, где $n \leq 0$, также являются функциями $e^{i\lambda n}$, $n \leq 0$, поэтому произведения конечных линейных комбинаций таких функций также являются конечными линейными комбинациями такого же вида. Переход к равномерному пределу приводит нас к функциям из $\mathbf{C}_{\leq 0}$. Пространства $\mathbf{C}_{\leq 0}$ и $\mathbf{C}_{\geq 0}$ полны относительно используемой для них нормы $\|f\|_\infty = \max_t |f(t)|$ (??5).

Нам понадобятся и другие аналогичные утверждения: i) произведение функции из $\mathbf{h}_{\leq 0}$ и функции из $\mathbf{C}_{\leq 0}$ принадлежит $\mathbf{h}_{\leq 0}$, ii) произведение функции из $\mathbf{h}_{>0}$ и функции из $\mathbf{C}_{\geq 0}$ принадлежит $\mathbf{h}_{>0}$. Оба утверждения проверяются одинаково, проверим i). Заметим (??6), что если $f_n \rightarrow f$ в $\mathbf{h}_{\leq 0}$, $g_n \rightarrow g$ в $\mathbf{C}_{\leq 0}$, то $f_n g_n \rightarrow fg$ в $\mathbf{h}_{\leq 0}$. Это доказывается элементарными оценками (??7).

Теперь мы займемся самой задачей прогноза на один шаг: прогноз на один шаг для процесса ξ_n по данным $\xi_n, n \leq 0$, обозначим через γ . Но мы процесс перевели в $L^2([-\pi, \pi], f_\xi(\lambda)d\lambda)$, в частности, $\xi_n, n \leq 0$, превратились в функции $e^{i\lambda n}, n \leq 0$, а функция, в которую превращается прогноз γ , мы обозначим g , разность значения процесса в момент 1 и прогноза для этого значения превращается в $e^{i\lambda} - g(\lambda)$. Итак,

$$e^{i\lambda} - g(\lambda) \perp e^{i\lambda n} \forall n \leq 0.$$

в пространстве $L^2([-\pi, \pi], f_\xi(\lambda)d\lambda)$. В интегралах это записывается следующим образом:

$$\int_{[-\pi, \pi)} [e^{i\lambda} - g(\lambda)] e^{i\lambda(-n)} f_\xi(\lambda) d\lambda = 0 \forall n \leq 0.$$

Теперь все это переведем в пространство $L^2([-\pi, \pi], d\lambda)$ (??8):

$$[e^{i\lambda} - g(\lambda)] f_\xi(\lambda) \in \mathbf{h}_{>0}, \quad (4)$$

$$g(\lambda) \in \mathbf{h}_{\leq 0}. \quad (5)$$

Далее мы будем решать задачу при выполнении условия *факторизации*:

$$f_\xi = f_1 f_2, f_1 \in \mathbf{C}_{\leq 0}, f_1^{-1} \in \mathbf{C}_{\leq 0}, f_2 = \bar{f}_1 \in \mathbf{C}_{\geq 0}, f_2^{-1} \in \mathbf{C}_{\geq 0}.$$

В результате (4) и благодаря утверждению i) (??9) свойство (4) заменяется на

$$h(\lambda) := [e^{i\lambda} - g(\lambda)] f_1 \in \mathbf{h}_{>0}. \quad (6)$$

Итак,

$$e^{i\lambda} f_1 = g(\lambda) f_1(\lambda) + h(\lambda), \text{ где } g \in \mathbf{h}_{\leq 0}, h \in \mathbf{h}_{>0}.$$

Более того, используя утверждение i) (??10) мы получаем

$$gf_1 \in \mathbf{h}_{\leq 0}, h \in \mathbf{h}_{>0}.$$

Таким образом, алгоритм нахождения наилучшего прогноза оказывается следующим (??11): надо взять функцию $e^{i\lambda} f_1$ и разложить ее в ряд Фурье по степеням $e^{in\lambda}$, $n \in \mathbf{Z}$. Часть ряда, принадлежащая $\mathbf{h}_{\leq 0}$, образует функцию gf_1 , разделив эту функцию на f_1 , получаем g .

Последнее, что надо сделать — показать, что спектральная плотность вида (3) удовлетворяет условию факторизации. Заметим, что из принятых выше соглашений следует, что числитель и знаменатель (3) не имеют корней на $[-\pi, \pi]$. Итак (??12), дробь (3) представляется в виде

$$f_\xi(\lambda) = C \frac{(z_1 - e^{i\lambda})(\bar{z}_1 - e^{-i\lambda}) \cdots (z_n - e^{i\lambda})(\bar{z}_n - e^{-i\lambda})}{(u_1 - e^{i\lambda})(\bar{u}_1 - e^{-i\lambda}) \cdots (u_m - e^{i\lambda})(\bar{u}_m - e^{-i\lambda})},$$

где C — положительная константа, z_k, u_j — корни соответствующих (??13) многочленов, не равные 1. Легко видеть, что все числа z_k, u_j можно считать по абсолютной величине строго больше 1 (??14 Где это потом используется?). Действительно, если $|z_k| < 1$, мы можем заменить соответствующие два множителя (изменяя при этом C), используя следующее равенство:

$$(z_k - e^{i\lambda})(\bar{z}_k - e^{-i\lambda}) = (z_k e^{-i\lambda} - 1)(\bar{z}_k e^{i\lambda} - 1) e^{i\lambda} e^{-i\lambda} = (e^{-i\lambda} - \frac{1}{z_k})(e^{i\lambda} - \frac{1}{\bar{z}_k}) |z^k|^2.$$

(??15. Как же так? Ведь согласно основной теореме алгебры, разложение на элементарные множители единственны.) Итак, мы считаем условие на абсолютные величины выполнеными и можем теперь выписать f_1 и f_2 :

$$f_1(\lambda) = \sqrt{C} \frac{(\bar{z}_1 - e^{-i\lambda}) \cdots (\bar{z}_n - e^{-i\lambda})}{(\bar{u}_1 - e^{-i\lambda}) \cdots (\bar{u}_m - e^{-i\lambda})}, f_2(\lambda) = \sqrt{C} \frac{(z_1 - e^{i\lambda}) \cdots (z_n - e^{i\lambda})}{(u_1 - e^{i\lambda}) \cdots (u_m - e^{i\lambda})}.$$

Числитель f_1 очевидным образом принадлежит $\mathbf{C}_{\leq 0}$, каждый множитель $\frac{1}{(u_m - e^{-i\lambda})}$ по формуле геометрической прогрессии разлагается в ряд по степеням $e^{-ik\lambda}$, $k \geq 0$, который сходится равномерно. Далее мы используем то, что $\mathbf{C}_{\leq 0}$ — алгебра.

Пример. $\xi_n = \zeta_n - 2\zeta_{n-1}$. Имеем: $f_\xi(\lambda) = \frac{1}{2\pi}(1 - 2e^{i\lambda})(1 - 2e^{-i\lambda})$, другое представление $f_\xi(\lambda) = \frac{2}{\pi}(1 - \frac{1}{2}e^{i\lambda})(1 - \frac{1}{2}e^{-i\lambda})$, $f_1(\lambda) = \frac{\sqrt{2}}{\sqrt{\pi}}(1 - \frac{1}{2}e^{-i\lambda})$, $e^{i\lambda} f_1(\lambda) = \frac{\sqrt{2}}{\sqrt{\pi}}(e^{i\lambda} - \frac{1}{2})$, $g(\lambda) f_1(\lambda) = \frac{\sqrt{2}}{\sqrt{\pi}}(-\frac{1}{2})$

$$g(\lambda) = \frac{\sqrt{2}}{\sqrt{\pi}} \left(-\frac{1}{2}\right) (f_1(\lambda))^{-1} = -\frac{1}{2} - \frac{1}{4}e^{-i\lambda} - \frac{1}{8}e^{-2i\lambda} \dots .$$

Оценка в момент времени 1 имеет вид:

$$\hat{\xi}_1 = -\frac{1}{2}\xi_0 - \frac{1}{4}\xi_{-1} - \frac{1}{8}\xi_{-2} \dots .$$

Добавление (не надо дальше учить).

Простейшая форма — некоторая определенная функция $f(x) + \eta_t$, где η_t — стационарный случайный процесс.

Процесс с комплексными значениями и ковариация такого стационарного процесса. Мотивация — процесс $\xi \cos(at)$ — заведомо нестационарный. Но если добавить $i\xi \sin(at)$, получится заведомо стационарный процесс ξe^{iat} . Если время дискретно, то $a \in [0, 2\pi]$. Задача оптимального выбора $f(x)$. Процесс белого шума в дискретном случае. Почему белый?

3. Безгранично делимые распределения

Определение 1. Безгранично делимое распределение вероятностей \mathbb{P} на числовой прямой — это такое распределение, что для любого натурального n существуют одинаково распределенные независимые случайные величины ξ_1, \dots, ξ_n , что

$$\mathbb{P} = \mathbb{P}_{\xi_1 + \dots + \xi_n}.$$

Представление Леви для характеристической функции φ произвольного безгранично делимого распределения:

$$\log \psi_{G,\gamma}(t) = i\gamma t + \int_{-\infty}^{\infty} \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) \frac{1+x^2}{x^2} dG(x) \quad (6)$$

где G — неубывающая ограниченная функция, γ — действительная постоянная, а подынтегральная функция в точке $x = 0$ считается равной $-t^2/2$.

Теорема. *Функция $\psi_{G,\gamma}$ является характеристической функцией безгранично делимого распределения.*

Доказательство. Так как

$$\psi_{G,\gamma} = (\psi_{G/n,\gamma/n})^n,$$

для доказательства безграничной делимости достаточно доказать, что $\psi_{G,\gamma}$ при любых G и γ является характеристической функцией распределения. Доказательство положительной определенности этой функции представляется довольно трудным, поэтому проще показать, что эта функция является пределом последовательности характеристических функций. Мы заменим интеграл (6) на верхнюю и нижнюю часть и каждую из них представим как предел интегралов по отрезку, а каждый из них — как предел интегральных сумм.

Итак,

$$\int_{\varepsilon}^{1/\varepsilon} \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) \frac{1+x^2}{x^2} dG(x) = \lim_n \sum_{k=0}^{n-1} T_{nk},$$

где

$$T_{nk} = \left(e^{it\xi_k} - 1 - \frac{it\xi_k}{1+\xi_k^2} \right) \frac{1+\xi_k^2}{\xi_k^2} (G(x_{k+1}) - G(x_k)),$$

$$\varepsilon = x_0 < \dots < x_n = 1/\varepsilon, x_k \leq \xi_k \leq x_{k+1}, \max_k |x_{k+1} - x_k| \rightarrow 0 (n \rightarrow \infty).$$

Каждое T_{nk} записывается в виде

$$T_{nk} = ita_{nk} + \lambda_{nk} (e^{itb_{nk}} - 1).$$

Соответственно, $e^{T_{nk}}$ является характеристической функции сдвинутого на a_{nk} распределения Пуассона с параметром λ_{nk} и шагом b_{nk} (а не 1, как у классического распределения Пуассона).

Определение 2. Последовательность $\{X_{nk} : k \leq k_n\}$ серий независимых случайных величин называется удовлетворяющей условию бесконечной малости, если

$$\max_{k \leq k_n} \mathbb{P}\{|X_{nk}| \geq \varepsilon\} \rightarrow 0 \text{ при } n \rightarrow \infty \text{ для любого фиксированного } \varepsilon > 0. \quad (7)$$

Теорема 2. Пусть последовательность $\{X_{nk} : k \leq k_n\}$ удовлетворяет условию бесконечной малости, τ — некоторое положительное число. Обозначим через F_{nk} функции распределения случайных величин X_{nk} ,

$$a_{nk} = \int_{|x|<\tau} x dF_{nk}(x); \bar{F}_{nk}(x) = F_{nk}(x + a_{nk});$$

$$\gamma_n = \sum_{k=1}^{k_n} \left\{ a_{nk} + \int_{-\infty}^{\infty} \frac{x}{1+x^2} d\bar{F}_{nk}(x) \right\};$$

$$G_{nk}(x) = \sum_{k=1}^{k_n} \int_{-\infty}^x \frac{y^2}{1+y^2} d\bar{F}_{nk}(y).$$

Тогда

$$\sum_{k=1}^{k_n} X_{nk} \xrightarrow{L} \varphi$$

тогда и только тогда, когда

$$G_n \xrightarrow{L} G, \gamma_n \rightarrow \gamma.$$

Доказательство. Обозначим характеристическую функцию X_{nk} через φ_{nk} , характеристическую функцию $\sum_{k=1}^{k_n} X_{nk}$ через φ_n , а распределение $\sum_{k=1}^{k_n} X_{nk}$ — через \mathbb{P}_n . Очевидная оценка

$$|\psi_{G,\gamma} - \varphi_n| \leq |\psi_{G,\gamma} - \psi_{G_n,\gamma_n}| + |\psi_{G_n,\gamma_n} - \varphi_n|$$

показывает, что доказательство сводится к доказательству двух сходимостей:

- a) $\psi_{G_n,\gamma_n}(t) \rightarrow \psi_{G,\gamma}(t)$ для всех $t \in \mathbf{R}$,
- b) $|\psi_{G_n,\gamma_n}(t) - \varphi_n(t)| \rightarrow 0$ для всех $t \in \mathbf{R}$.

Первое утверждение является самостоятельным фактом — теоремой о сходимости характеристических функций. Впрочем, для полноты доказательства нам нужно доказать еще слабую компактность семейства распределений $\{\mathbb{P}_n\}$.

НЕОБХОДИМЫЕ СВЕДЕНИЯ ИЗ ДРУГИХ МАТЕМАТИЧЕСКИХ КУРСОВ

Анализ

Производная сложной функции

Критерии Коши сходимости последовательности и ряда

Критерий ε -милиционеров

Критерий сходимости ряда в терминах абсолютно сходящегося ряда

Равномерная непрерывность и теорема Кантора

Компактность отрезка на числовой прямой

Интегральные суммы Римана — Стильбеса

Преобразование Фурье

Линейная алгебра

Дополнение ортонормированной системы до ортонормированного базиса

Ортогональная матрица

Неотрицательно определенная квадратичная форма

Сумма квадратов как произведение вектора-строки на вектор-столбец

Вращение и сумма квадратов координат

Теория меры

Продолжение меры с полуалгебры на алгебры, с алгебры на σ -алгебру

Эквивалентность счетной аддитивности и непрерывности для конечно аддитивной функции множеств на алгебре

Определение интеграла Лебега по вероятностной мере

Теорема Лебега для сходимости по вероятности

Теорема Радона — Никодима

Функциональный анализ

Теорема Банаха о неподвижной точке

Теорема Брауэра о неподвижной точке

Комплексный анализ

Интеграл по замкнутому контуру

Дифференциальные уравнения

Метод вариации постоянной.

Уравнения с частными производными.

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

по курсам 'Теория вероятностей', 'Математическая статистика',
'Случайные процессы'
для специальности 'Математика'

Учебники

1. А.А. Боровков. Теория вероятностей. Наука, 1976.
2. А.А. Боровков. Математическая статистика. Оценка параметров. Проверка гипотез. Наука, 1984.
3. А.В. Булинский, А.Н. Ширяев. Теория случайных процессов. Физматлит, 2003.
4. А.Д. Вентцель. Курс теории случайных процессов. Наука, 1975.
5. И.Н. Володин. Лекции по теории вероятностей и математической статистике. Изд-во КГУ. 2006. 271 с.
6. И.Н. Володин, О.Е. Тихонов, Е.А. Турилова. Математические основы вероятности. Изд-во КГУ, 2006. 163 с.
7. Б.В. Гнеденко. Курс теории вероятностей. М. Наука, 1969.
8. И.И. Гихман, А.В. Скороход. Введение в теорию случайных процессов. Наука. 1967. 567 с.
9. Г.И. Ивченко, Ю.И. Медведев. Математическая статистика. Изд-во Высшая школа. 1984, 248 с.
10. Г.П. Климов. Теория вероятностей и математическая статистика. Изд-во МГУ. 1983. 328 с.
11. М.В. Козлов, А.В. Прохоров. Введение в математическую статистику. Изд. МГУ, 1987.
12. Г.Крамер. Математические методы статистики. Мир, 1976.
13. М.Б. Лагутин. Наглядная математическая статистика. М.: БИНОМ. Лаборатория знаний. 2007. 472 с.
14. Методические разработки по специальному курсу 'Многомерный статистический анализ' (составитель С.В. Симушкин). Изд-во КГУ. 2006. 96 с.
15. Д.Х. Муштари. Вероятность, математическая статистика, случайные процессы. Электронное учебное пособие. Site KGU: <http://www.ksu.ru>. 188 с.
16. Ю.А. Розанов. Теория вероятностей, случайные процессы и математическая статистика. Главная редакция физико-математической литературы, 1985.
17. Б.А. Севастьянов. Курс теории вероятностей и математической статистики. М. Наука, 1982.
18. В.Н. Тутубалин. Теория вероятностей. Изд-во МГУ, 1972.
19. А.Н..Ширяев. Вероятность. Наука, 1980.
20. R. Bhattacharya, E.C. Waymire. A Basic Course in Probability theory. Springer. Universitext. 2007. ix+211.
21. L.C. Evans. An Introduction to Stochasstic Differential Equations. Version I,2. University Berkeley. 139 p.
22. Jun Shao. Mathematical Statistics. Springer. 2003. xvi+591.
23. D. Kannan. An Introduction to Stochastic Processes. North Holland, 1979. xiii + 296 p.
24. O. Knill. Probability. 1994. 227 р. (электронный курс лекций для студентов Калифорнийского технологического института).
25. L.B. Koralov, Ya.G. Sinai. Theory of Probability and Random Processes. Springer. Universitext. 2007. xi + 353 p.

Задачники

26. А.М. Зубков. Сборник задач по теории вероятностей. Изд. МГУ.
27. Л.Д. Мешалкин. Сборник задач по теории вероятностей. Изд. МГУ. 1963.

28. Практические занятия по теории вероятностей (И.И. Адгамов, Д.Х. Муштари). Методическое пособие. Изд. КГУ, 1989. 38 с.
29. А.В. Прохоров, В.Г. Ушаков, Н.Г. Ушаков. Задачи по теории вероятностей. Наука. 1986.
30. Б.А. Севастьянов, В.П. Чистяков, А.М. Зубков. Сборник задач по теории вероятностей. Наука. 1983.
31. А.В. Сульдин, Е.А. Беговатов, С.В. Григорьев. Сборник задач по теории вероятностей и математической статистике. Изд. Казанского университета. Казань. 1980. 94 с.
32. Д.М. Чибисов, В.И. Пагурова. Задачи по математической статистике. Изд. Московского университета. 1990. 171 с.
33. Jun Shao. Mathematical Statistics: Exercises and Solutions. Springer. 2005.xxviii+359.

Справочники

34. Л.Н. Больщев, Н.В. Смирнов. Таблицы математической статистики. М. Наука, 1983.
35. Энциклопедия 'Вероятность и математическая статистика' (Гл. ред. – Ю.В. Прохоров). М. Большая Российская Энциклопедия, 1999.

Монографии

36. Г.П. Климов. Инвариантные выводы в статистике. Изд-во МГУ. 1973. 186 с.
37. Д. Лоули. А. Максвелл. Факторный анализ как статистический метод. Библиотека сборника 'Математика'. Изд. 'Мир'. М. 1967. 144 с.
38. А.Н. Ширяев. Основы стохастической финансовой математики. Том 1. Факты. Модели, Том 2. Теория. Изд. 'Фазис'. 1998.

ЭКЗАМЕНЫ И ЗАЧЕТЫ

Экзамен по теории вероятностей

ПРОГРАММА

Частотная интерпретация вероятности. Вероятностное пространство. Теория вероятностей и теория меры (сравнение обозначений и терминологий).

СОБЫТИЯ. Классическая модель. Различные варианты выборок. Задача о k гербах в n экспериментах. Модели Больцмана — Максвелла (полиномиальная) и Бозе — Эйнштейна. Геометрическая модель.

Формула сложения. Условная вероятность и независимость. Формула умножения. Задача о распаде атома. Формулы полной вероятности и Байеса. Пример на контроль качества. Задача о разорении игрока.

Независимость. Условная независимость. Эквивалентные определения независимости в совокупности. Операции с независимыми событиями. Примеры: электрическая цепь и пример Бернштейна.

Модель Бернулли. Предельная теорема Пуассона.

Независимые полуалгебры и алгебры. Связь счетной аддитивности и непрерывности. Независимость σ -алгебр и теорема о монотонных классах.

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ. Случайные величины и их распределения на борелевской σ -алгебре. Два определения случайной величины (прообразы борелевских множеств и прообразы интервалов), их эквивалентность. Борелевская функция случайной величины — случайная величина. Свойства функции распределения (в частности, непрерывность вероятностной меры). Примеры распределений (Бернулли, биномиальное, Пуассона, нормальное, равномерное, экспоненциальное). Вывод плотности нормального распределения с произвольными параметрами. Связь функции распределения и функции плотности. Распределения квадрата и экспоненты.

Случайные векторы и наборы случайных величин. Распределения случайных векторов (совместные распределения случайных величин). Связь совместной функции распределения и совместной функции плотности. Задача о плотности распределения проекции случайного вектора. Задание распределения по совместной функции распределения.

Независимые случайные величины. Эквивалентность двух определений независимости случайных величин. (в частности, операции с независимыми событиями, сохраняющие независимость). Функции нескольких независимых случайных величин и теорема о монотонных классах. Независимость на языке совместной плотности для непрерывного случая.

Распределения, связанные с набором независимых нормальных случайных величин: Коши, хи-квадрат (с выводом), Стьюдента.

Многомерное нормальное распределение (вводится только невырожденное). Стандартное нормальное распределение.

Лемма о вращении стандартного нормального случайного вектора.

Среднее и его свойства. Доказательства (Формула замены переменной. Среднее произведение независимых случайных величин). Вычисление среднего в случае дискретного или непрерывного распределений. Дисперсия и ее свойства. Вычисление среднего и дисперсии для некоторых распределений (биномиальное, нормальное, равномерное, Пуассона, χ^2 , Стьюдента). Пример отсутствия среднего. Моменты и другие характеристики.

Неотрицательная определенность матрицы вторых моментов. Коэффициент корреляции и его свойства. Связь с независимостью. Случай совместного нормального распределения. Пример двух зависимых симметричных случайных величин со значениями в $\{0, +1, -1\}$ с нулевым коэффициентом корреляции. Функция регрессии (случай константы, дискретной случайной величины, совместной функции плотности и нормального распределения).

ПОСЛЕДОВАТЕЛЬНОСТИ СЛУЧАЙНЫХ ВЕЛИЧИН. Сходимость по вероятности. Неравенство Чебышева. Слабый закон больших чисел. Применение к частоте события. Метод Монте-Карло.

Лемма Бореля — Кантелли. Критерий Коши сходимости ряда почти наверное. Неравенство

Колмогорова. Сходимость по Чезаро и связь с обычной сходимостью. Усиленный закон больших чисел для разно распределенных случайных величин со вторым моментом. Усиленный закон больших чисел Колмогорова для одинаково распределенных случайных величин.

Слабая сходимость распределений. Эквивалентность двух определений (на языке интегралов и функций распределений). Связь сходимости по вероятности и слабой сходимости.

ЗАДАЧИ ИЗ ПРАКТИКИ (не включается среднее значение, дисперсия и т.д.), будут две задачи из всего предшествующего (события, формулы, независимость, случайные величины, случайные векторы, их распределения и распределения функций, в том числе независимые случайные величины).

Вопросы, на которые нужно отвечать сразу:

1) Определение независимых событий. Нужно понимать, что верно и обратное, если вероятность произведения равна произведению вероятностей, то эти события независимы, даже если на первый взгляд кажутся зависимыми. Верно и противоположное, если нужного равенства вероятностей нет, то события зависимы.

2) События, независимые в совокупности. Вероятность произведения равна произведению вероятностей для любого поднабора.

3) Несовместные события, их нельзя путать с независимыми. Нужно понимать, что для несовместных A и B имеет место $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ (мы пишем, используя дополнительную нагрузку на символ, $\mathbb{P}(A + B) = \mathbb{P}(A) + \mathbb{P}(B)$). Это частный случай формулы сложения, используется при решении задач. Для независимых событий верна другая формула (ее помнить не обязательно) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)$.

Итак, формула $\mathbb{P}(A + B) = \mathbb{P}(A) + \mathbb{P}(B)$ будет использоваться в самых разных вариантах. Например, $\mathbb{P}(A) = \mathbb{P}(\Omega) - \mathbb{P}(\bar{A})$ — частный случай формулы $\mathbb{P}(A) = \mathbb{P}(A + B) - \mathbb{P}(B)$. Итак, для вычисления (или оценки сверху) вероятности события его нужно разбить на несколько несовместных событий, каждое из которых легче вычислить (или оценить).

4) Определение условной вероятности.

5) Вычисление вероятности попадания случайного вектора в множество в дискретном и непрерывном случаях. Приведем алгоритм для двумерного случайного вектора (ξ, η) . В дискретном случае знание распределения означает знание всех вероятностей

$$\mathbb{P}\{(\xi, \eta) = (x_i, y_i)\} = p_i, \text{ где } \sum_i p_i = 1.$$

Итак, для вычисления $\mathbb{P}\{(\xi, \eta)\}$, надо сложить все числа p_i , для которых $(x_i, y_i) \in B$. Например, для вычисления $\mathbb{P}\{\xi = x\}$ надо сложить все числа p_i , для которых $x_i = x$ (в этом случае $B = \{x\} \times \mathbf{R}$).

Аналогично, в непрерывном случае мы должны взять интеграл по B от совместной функции плотности. Обратно, это равенство является определением совместной функции плотности. Поэтому в двумерном случае функция плотности ξ задается интегралом (точнее, может быть задана интегралом, так как функция плотности задана с точностью до значений на множестве лебеговой меры нуль)

$$p_\xi(x) = \int_{\mathbf{R}} p_{\xi, \eta}(x, y) dy.$$

Действительно, при таком задании мы будем иметь тождество для любого борелевского B :

$$\begin{aligned} \mathbb{P}\{\xi \in B\} &= \int_B p_\xi(x) dx = \\ &= \mathbb{P}\{(\xi, \eta) \in B \times \mathbf{R}\} = \int_{B \times \mathbf{R}} p_{\xi, \eta}(x, y) dx dy. \end{aligned}$$

6) Определение независимости случайных величин: ξ и η :

$$\mathbb{P}\{(\xi, \eta) \in B_1 \times B_2\} = \mathbb{P}\{\xi \in B_1\} \cdot \mathbb{P}\{\eta \in B_2\}.$$

В дискретном случае нужно проверить, что для всех возможных значений x и y случайных величин ξ и η имеет место равенство

$$\mathbb{P}\{\xi = x, \eta = y\} = \mathbb{P}\{\xi = x\} \cdot \mathbb{P}\{\eta = y\}. \quad (1)$$

Определение получается сложением этих равенств по всем возможным значениям $x \in B_1$ и $y \in B_2$ случайных величин ξ и η . В непрерывном случае для независимости необходимо и достаточно разложение

$$p_{\xi, \eta}(x, y) = p_\xi(x) \cdot p_\eta(y) \quad (2)$$

для всех x и y , за исключением множества лебеговой меры нуль.

7) Как записывается совместная функция плотности независимых случайных величин $\xi_1, \xi_2, \dots, \xi_n$, имеющих одну и ту же функцию плотности p .

8) Вычисление среднего в дискретном случае

$$\mathbb{E}\xi = \sum_i x_i \mathbb{P}\{\xi = x_i\},$$

суммирование по всем возможным значениям x_i случайной величины ξ . Соответственно, если заданы вероятности значений случайного вектора (ξ, η) , то

$$\mathbb{E}\xi = \sum_i x_i \mathbb{P}\{\xi = x_i, \eta = y_i\},$$

суммирование по всем возможным значениям (x_i, y_i) случайного вектора (ξ, η) .

9) Вычисление среднего в непрерывном случае

$$\mathbb{E}\xi = \int_{\mathbf{R}} xp_\xi(x) dx.$$

Соответственно, если задана совместная плотность случайного вектора (ξ, η) , то

$$\mathbb{E}\xi = \int_{\mathbf{R}} \int_{\mathbf{R}} xp_{\xi, \eta}(x, y) dx dy.$$

10) Аддитивность среднего, константа выносится из под знака среднего, среднее константы равно этой константе, в частности $\mathbb{E}(\mathbb{E}\xi) = \mathbb{E}\xi$ (среднее значение — это константа).

11) Среднее произведения независимых случайных величин равно произведению средних.

12) Два определения дисперсии:

$$\mathbb{D}\xi = \mathbb{E}(\xi - \mathbb{E}\xi)^2 = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2.$$

13. Свойства дисперсии — дисперсия не меняется при сдвиге случайной величины на константу ($\mathbb{D}\xi = \mathbb{D}(\xi + c)$), при умножении случайной величины на константу ее дисперсия умножается на квадрат константы, и, наконец, дисперсия суммы независимых случайных величин равна сумме дисперсий.

14) Параметры нормального распределения: m — среднее, σ^2 — дисперсия.

15) Определение функции распределения, ее связь с функцией плотности.

Требование отвечать на эти вопросы сразу вынужденно. Все это многократно используется, и если не воспринимать эти вещи как очевидные, то вы не будете успевать понимать лекции в следующем семестре. Так что рекомендуется несколько раз все это повторить, причем в конце ваших занятий, когда вы устали. Спрашиваться это будет в конце экзамена.

Экзамен по математической статистике

Кроме материала курса математической статистики надо на уровне определений и формулировок знать следующий материал:

Квадратичная форма. Вращение квадратичной формы. Ортогональная матрица, связь с ортонормированным базисом. Дополнение ортонормированной системы до ортонормированного базиса. Независимые события. Условная вероятность, формула полной вероятности. Формула полной вероятности для плотности и условной плотности. Распределение Бернулли. Распределение Пуассона. Нормальное распределение. Экспоненциальное распределение. Равномерное распределение. Лемма о вращении стандартного нормального случайного вектора. Независимые случайные величины. Свойства функций независимых случайных величин. Совместная плотность независимых случайных величин. Совместная плотность одинаково распределенных независимых случайных величин. Среднее, дисперсия, коэффициент корреляции. Алгебраические свойства среднего и дисперсии. Две формулы для дисперсии. Сходимость по вероятности. Закон больших чисел. Слабая сходимость распределений, эквивалентность двух определений. Где в курсе математической статистики используются все эти понятия и результаты? (Весь раздел без доказательства)

Программа

Характеристическая функция. Элементарные свойства характеристической функции. Формула обращения. Теорема единственности.

Теорема о слабой компактности. Критерий слабой компактности в пространстве функций распределения. Неравенство для усечений. Доказательство теоремы непрерывности.

Вычисление характеристической функции нормального распределения.

Применения характеристических функций. Центральная предельная теорема для одинаково распределенных слагаемых. Центральная предельная теорема в схеме серий (теорема Линдеберга). Аппроксимация распределения χ^2 и Стьюдента нормальным распределением. Моделирование нормального распределения

Многомерная характеристическая функция, элементарные свойства и многомерные теорема единственности и непрерывности (без доказательств). Определение многомерного нормального распределения общего вида (возможно, вырожденного). Многомерная центральная предельная теорема.

Эмпирическое распределение и его характеристики. Вероятностные и статистические модели. Примеры моделей. График эмпирической функции распределения. Оценивание параметров. Состоительные оценки. Несмешенные оценки. Несмешенная оценка среднего и дисперсии в нормальной модели.

Эффективные оценки. Неравенство Рао — Крамера. Теорема о том, что эффективная оценка является оценкой максимального правдоподобия. Алгоритм нахождения эффективной оценки (если она существует). Эффективная оценка среднего в нормальной модели. Оценка дисперсии в нормальной модели. Контрпример для эмпирического среднего (как оценки среднего равномерного распределения).

Состоительная оценка. Примеры и контрпример для модели Коши. Теорема о состоятельности оценки максимального правдоподобия (без доказательства неравенства Йенсена). Пример.

Доверительные интервалы. Построение доверительного интервала для среднего (при неизвестной дисперсии) и для дисперсии (при неизвестном среднем). Связь понятия состоятельной оценки и понятия доверительного интервала.

Достаточные статистики и теорема факторизации.

Проверка статистических гипотез. Критерий Стьюдента (условия применения и вывод). Критерий χ^2 . Вывод предельного распределения статистики χ^2 (с использованием многомерной центральной предельной теоремы). Оценка минимума χ^2 . Случай нескольких параметров. Применение для проверки независимости двух признаков. Критерий Колмогорова — Смирнова и сведение вычисления уровня значимости к случаю равномерного распределения. Критерий знаков. Эмпирический коэффициент корреляции и вывод его распределения в предположении независимости в нормальной модели. Проверка независимости в нормальной модели.

Ранговые критерии (понятие). Критерий однородности, критерий серий. Критерий Вилкоксона. Асимметрия и эксцесс. Проверки на нормальность выборки. Критерий Фишера.

Задача сравнения гипотез. Лемма Неймана – Пирсона. Равномерно наиболее мощные критерии. Примеры существования и несуществования равномерно наиболее мощных критериев. Область безразличия.

Последовательный критерий Вальда.

Последние вопросы прочитать по данному электронному учебному пособию. Регрессия. Корреляция. Факторный анализ. Метод главных компонент.

В зависимости от посещения практических занятий может быть предложен вопрос построения доверительных интервалов или проверки гипотезы на конкретных числах. Если в программе имеются вопросы, которых не было на лекциях (за исключением последнего раздела), то на основании своих (!) лекций это можно доказать.

Зачет по теории случайных процессов

Кроме материала курса теории случайных процессов надо на уровне определений и формулировок знать следующий материал:

Независимые события. Условная вероятность, формула полной вероятности. Распределение Бернулли. Распределение Пуассона. Нормальное распределение. Экспоненциальное распределение. Независимые случайные величины. Свойства функций независимых случайных величин. Совместная плотность независимых случайных величин. Совместная плотность одинаково распределенных независимых случайных величин. Среднее, дисперсия. Алгебраические свойства среднего и дисперсии. Две формулы для дисперсии. Сходимость по вероятности. Закон больших чисел. Характеристическая функция. Формула обращения. Теорема единственности. Элементарные свойства характеристической функции. Вычисление характеристической функции нормального распределения. Центральная предельная теорема для одинаково распределенных слагаемых. (Весь раздел без доказательства). Где в курсе теории случайных процессов используются все эти понятия и результаты?

Основные понятия теории случайных процессов (отвечать сразу определения и формулировки)

1. Что такое случайный процесс с непрерывным и дискретным временем.
2. Среднее процесса и ковариация процесса
3. Совместные распределения процесса
4. Свойства траекторий процесса
5. Аксиомы винеровского процесса. Одномерные распределения винеровского процесса со сносом и диффузией
6. Аксиомы пуассоновского процесса. Одномерные распределения пуассоновского процесса
7. Что такое стационарный случайный процесс
8. В чем состоит линейный прогноз стационарного процесса. На что происходит проектирование
9. Что такая спектральная плотность стационарного случайного процесса.
10. Что такое стохастический интеграл Ито, какие у него интегральные суммы?
11. Что такое мартингал?
12. Что такое процесс случайного блуждания?
13. Что такое однородная цепь Маркова?

Программа

Лемма Бореля — Кантелли. Критерий Коши сходимости ряда почти наверное. Неравенство Колмогорова. Сходимость по Чезаро и связь с обычной сходимостью. Усиленный закон больших чисел для разно распределенных случайных величин со вторым моментом. Усиленный закон больших чисел Колмогорова для одинаково распределенных случайных величин.

Одномерное случайное блуждание. Задача о пьяном гуляке и принцип отражения. Задача о баллотировке. Марковское свойство случайного блуждания. Задача о постоянном везении. Задача о возвращении случайного блуждания в начало на прямой, на плоскости, в пространстве. Задача о продолжительном везении. Распределение арксинуса.

Марковский процесс и цепь Маркова. Однородная цепь Маркова. Переходные вероятности. Задача о возвращении для однородной цепи Маркова со счетным числом состояний. Теорема о существовании инвариантного состояния у цепи Маркова.

Пуассоновский и винеровский процессы: параметры, вывод распределений, ковариация, совместное распределение. Производящая функция. Процессы гибели и размножения.

Оценка параметров сноса и диффузии процесса Винера. Теорема Колмогорова о непрерывности траекторий. Непрерывность траекторий процесса Винера. Применение — сходимость почти наверное к коэффициенту сноса процесса. Недифференцируемость траекторий винеровского процесса.

Марковские процессы. Диффузионные процессы.

Стационарные процессы. Прогноз для стационарного процесса. Процессы скользящего среднего и авторегрессии. Процесс Орнштейна — Уленбека.

Интеграл Ито. Формула Ито. Два вывода $\int_0^T w_t dw_t$. Пример применения формулы Ито для решения одного стохастического дифференциального уравнения.

Заряды. Теоремы Хана и Радона — Никодима. Понятие об условном среднем. Условная вероятность относительно σ -алгебры. Мартингалы. Теорема Дуба о почти наверное сходимости мартингала. Мартингалы в финансовой математике. Задача об оптимальном моменте остановки. Применение: задача о разборчивой невесте.

ТАБЛИЦЫ ВЕРОЯТНОСТЕЙ (см. [30])

Таблица 1. Функция распределения Φ стандартного нормального распределения:

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Горизонтальная часть таблицы указывает сотые доли x . Значения Φ умножены на 10^4 .

x	0	1	2	3	4	5	6	7	8	9
0.0	5000	5040	5080	5120	5160	5200	5239	5279	5319	5359
0.1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0.2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0.3	6179	6217	6256	693	6331	6338	6406	6443	6480	6517
0.4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0.5	6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0.6	7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0.7	7580	7611	7642	7673	7703	7734	7764	7794	7823	7852
0.8	7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
0.9	8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1.0	8413	8437	8461	8485	8508	8381	8554	8577	8599	8621
1.1	864	8665	8686	8708	8729	8749	8770	8790	8810	8831
1.2	8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
1.3	9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
1.4	9192	9207	9222	9236	9251	9255	9279	9292	9306	9319
1.5	9332	9345	9357	9370	9382	9394	9406	9418	9429	9441
1.6	9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
1.7	9554	9564	9573	9582	8591	9599	9608	9616	9625	9633
1.8	9641	9649	9656	9664	9671	9678	9686	9693	9699	9705
1.9	9713	9719	9726	9732	9738	9744	9750	9756	9761	9767
2.0	9772	9778	9783	9788	9793	9798	9803	9808	9812	9817
2.1	9821	9825	9830	9834	9838	9842	9846	9850	9854	9857
2.2	9861	9864	9868	9871	9875	9878	9881	9884	9887	9890
2.3	9893	9896	9898	9901	9904	9906	9909	9911	9913	9915
2.4	9918	9920	9922	9925	9927	9929	9931	9932	9934	9936
2.5	9938	9940	9941	9943	9945	9946	9948	9949	9951	9952
2.6	9953	9955	9956	9957	9959	9960	9961	9962	9963	9964
2.7	9965	9966	9967	9968	9969	9970	9971	9972	9973	9974
2.8	9974	9975	9976	9977	9977	9978	9979	9979	9980	9981
2.9	9981	9982	9982	9983	9984	9984	9985	9985	9986	9986
3.0	9987	9987	9987	9988	9988	9989	9989	9989	9990	9990

Таблица 2. $(1 - p)$ -квантиль q распределения χ^2 : $\mathbb{P}\{\chi_n^2 > q\} = p$.

$n \setminus p$	0.995	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
2	0.0100	0.0201	0.0506	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.0717	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.336	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.688	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
31	14.458	15.655	17.539	19.281	21.434	41.422	44.985	48.232	52.191	55.003
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328
33	15.815	17.073	19.047	20.867	23.110	43.745	47.400	50.725	54.776	57.648
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964
35	17.192	18.509	20.569	22.465	24.797	46.059	49.802	53.203	57.342	60.275
36	17.887	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581
37	18.586	19.960	22.106	24.075	26.492	48.363	52.192	55.668	59.892	62.882
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.895	61.162	64.181
39	19.996	21.426	23.654	25.695	28.196	50.660	54, 572	58.120	62.428	65.476
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766

Таблица 3. $(1 - p)$ -квантиль q распределения Стьюдента случайной величины t_n : $\mathbb{P}\{t_n > q\} = p$.

$n \setminus p$	0.05	0.025	0.01	0.005	$n \setminus p$	0.05	0.025	0.01	0.005
1	6.3138	12.7062	31.8205	63.6567	28	1.7011	2.0484	2.4671	2.7633
2	2.9200	4.3027	6.9646	9.9248	29	6991	0452	4620	7564
3	3534	3.1824	4.5407	5.8409	30	6973	0423	4573	7500
4	1318	2.7764	3.7469	4.6041	32	6939	0369	4487	7385
5	4759	5706	3649	0321	34	6909	0322	4411	7284
6	1.9432	2.4469	3.1427	3.7074	36	6883	0281	4345	7195
7	8946	3646	2.9980	4995	38	6860	0244	4286	7116
8	8595	3060	8965	3554	40	6839	0211	4233	7045
9	8331	2622	8214	2498	42	1.6820	2.0181	2.4185	2.6981
10	8125	2281	7638	1693	44	6802	0154	4141	6923
11	1.7959	2.2010	2.7181	3.1058	46	6787	0129	4102	6870
12	7823	1788	6810	0545	48	6772	0106	4066	6822
13	7709	1604	6503	0123	50	6759	0086	4033	6778
14	7613	1448	6245	2.9768	55	2.6730	2.0040	2.3961	2.6682
15	7530	1314	6025	9467	60	6706	0003	3901	6603
16	1.7459	2.1199	2.5835	2.9208	65	6686	1.9971	3851	6536
17	7396	1098	5669	8982	70	6669	9944	3808	6479
18	7341	1009	5524	8784	80	6641	9901	3739	6387
19	7291	0930	5395	8609	90	6620	9867	3685	6316
20	7247	0860	5280	8453	100	1.6602	1.9840	2.3642	2.6259
21	1.7207	2.0796	2.5176	2.8314	120	6577	9799	3578	6174
22	7171	0739	5083	8188	150	6551	9759	3515	6090
23	7139	0687	4999	8073	200	6525	9719	3451	6006
24	7109	0639	4922	7969	250	1.6510	1.9695	2.3414	2.5956
25	7081	0595	4851	7874	300	6499	9679	3388	5923
26	7056	0555	4786	7787	400	6487	9659	3357	5882
27	1.7033	2.0518	2.4727	2.7707	500	1.6479	1.9647	2.3338	2.5857

Таблица 4. 0.95-квантиль q распределения Фишера случайной величины

$$F_{m,n} = \frac{n\chi_m^2}{m\chi_n^2}, \mathbb{P}\{F_{m,n} > q\} = 0.05.$$

$n \setminus m$	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8868	8.8452	8.8123
4	7.7086	6.9443	6.5914	6.3883	6.2560	6.1631	6.0942	6.0410	5.9988
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2066	4.1468	4.0990
7	5.5914	4.7374	4.3468	4.1203	4.9715	3.8660	3.7870	3.7257	3.6767
8	5.3177	4.4590	4.0662	3.8378	3.6875	3.5806	3.5005	3.4381	3.3881
9	5.1174	4.2565	3.8626	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377

Таблицы, используемые при проверке выборки на нормальность

В следующих таблицах для выборки (x_1, \dots, x_n) используются стандартные обозначения:

$$\bar{x} = \frac{x_1 + \dots + x_n}{n}, \quad s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Таблица 5. $(1-p)$ -квантиль q распределения выборочной характеристики эксцесса в нормальной модели с выборкой (x_1, \dots, x_n)

$$\text{Эксцесс обозначается в [30]} \quad b_2 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{s^4}.$$

Число в таблице q определяется равенством $\mathbb{P}\{b_2 > q\} = p$.

$n \setminus p$	0.01	0.05	0.95	0.99
50	4.92	4.01	2.13	1.95
100	4.40	3.77	2.35	2.18
150	4.14	3.66	2.45	2.30
200	3.98	3.57	2.51	2.37
250	3.87	3.51	2.55	2.42

Таблица 6. $(1-p)$ -квантиль q распределения выборочной характеристики асимметрии g_1 в нормальной модели с выборкой (x_1, \dots, x_n)

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$$

$$\text{Асимметрия обозначается в [30]} \quad g_1 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{s^3}.$$

Число в таблице q определяется равенством $\mathbb{P}\{g_1 > q\} = p$.

$n \setminus p$	0.05	0.01	$\sqrt{\mathbb{D}g_1}$
25	0.711	1.061	0.4354
30	0.661	0.982	0.4052
35	0.621	0.921	0.3804
40	0.587	0.869	0.3596
45	0.558	0.825	0.3418
50	0.533	0.787	0.3264
60	0.492	0.723	0.3009
70	0.459	0.673	0.2806
80	0.432	0.631	0.2638
90	0.409	0.596	0.2498
100	0.389	0.567	0.2377

Таблица 7. $(1 - p)$ -квантиль q распределения выборочной характеристики d в нормальной модели с выборкой (x_1, \dots, x_n)

$$d = \frac{\frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|}{s}.$$

Число в таблице q определяется равенством $\mathbb{P}\{d > q\} = p$.

$n \setminus p$	0.01	0.05	0.1	0.9	0.95	0.99	$\sqrt{\mathbb{D}d}$
11	0.9359	0.9073	0.8899	0.7409	0.7153	0.6675	0.05784
16	9137	8884	8733	7452	7236	6829	0.04976
21	9001	8768	8631	7495	7304	6950	0.04419
26	8901	8686	8570	7530	7360	7040	0.04011
31	8827	8625	8511	7559	7404	7110	0.03697
36	8769	8578	8468	7583	7440	7167	0.03447
41	8722	8540	8436	7604	7470	7216	0.03241
46	8682	8508	8409	7621	7496	7256	0.03068
51	8648	8481	8385	7636	7518	7291	0.02919
61	0.8592	0.8434	0.8349	0.7662	0.7554	0.7347	0.02678
71	8549	8403	8321	7683	7583	7393	0.02487
81	8515	8376	8298	7700	7607	7430	0.02332
91	8484	8353	8279	7714	7626	7460	0.02203
101	8460	8344	8264	7726	7644	7487	0.02094

Примечание. Для таблиц 6 и 7 для n , отличных от использованных в таблице, в [30] предлагаются использовать линейную интерполяцию или экстраполяцию. Причем интерполяция или экстраполяция должны проводиться не по аргументу n , а по аргументу $\sqrt{\mathbb{D}g_1}$ в таблице 6 и по аргументу $\sqrt{\mathbb{D}d}$ в таблице 7. Там же приведены формулы для дисперсий:

$$\mathbb{D}g_1 = \frac{6(n-2)}{(n+1)(n+3)}$$

$$Dd = \frac{1}{n} \left\{ 1 + \frac{2}{\pi} \left[\sqrt{n(n-2)} + \arcsin \frac{1}{n-1} \right] \right\} - \frac{n-1}{\pi} \left[\frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})} \right]^2.$$

Таблица 8. Функция распределения рангового коэффициента корреляции Спирмена ρ в нормальной модели с выборкой (x_1, \dots, x_n)

В таблице для каждого числа наблюдений n указаны вероятности $p(n) = \mathbb{P}\{\rho \leq r(n)\}$, где числа $r(n)$ пробегают значения (не все), которые для данного n может принимать ρ .

$r(4)$	$p(4)$	$r(5)$	$p(5)$	$r(6)$	$p(6)$	$r(7)$	$p(7)$	$r(8)$	$p(8)$	$r(9)$	$p(9)$	$r(10)$	$p(10)$
$-\frac{1}{5}$	0.458	$-\frac{12}{120}$	0.475	$-\frac{90}{210}$	0.210	$-\frac{116}{336}$	0.249	$-\frac{144}{504}$	0.250	$-\frac{216}{720}$	0.218	$-\frac{258}{990}$	0.235
$-\frac{2}{5}$	0.375	$-\frac{24}{120}$	0.392	$-\frac{102}{210}$	0.178	$-\frac{128}{336}$	0.198	$-\frac{180}{504}$	0.195	$-\frac{264}{720}$	0.168	$-\frac{318}{990}$	0.184
$-\frac{3}{5}$	0.208	$-\frac{36}{120}$	0.342	$-\frac{114}{210}$	0.149	$-\frac{140}{336}$	0.151	$-\frac{216}{504}$	0.150	$-\frac{312}{720}$	0.125	$-\frac{378}{990}$	0.139
$-\frac{4}{5}$	0.167	$-\frac{48}{120}$	0.258	$-\frac{126}{210}$	0.121	$-\frac{152}{336}$	0.118	$-\frac{252}{504}$	0.108	$-\frac{360}{720}$	0.089	$-\frac{438}{990}$	0.102
-1	0.042	$-\frac{60}{120}$	0.225	$-\frac{138}{210}$	0.088	$-\frac{164}{336}$	0.083	$-\frac{288}{504}$	0.076	$-\frac{408}{720}$	0.060	$-\frac{498}{990}$	0.072
		$-\frac{72}{120}$	0, 175	$-\frac{150}{210}$	0.068	$-\frac{176}{336}$	0.055	$-\frac{324}{504}$	0.048	$-\frac{456}{720}$	0.038	$-\frac{558}{990}$	0.048
		$-\frac{84}{120}$	0.117	$-\frac{162}{210}$	0.051	$-\frac{188}{336}$	0.033	$-\frac{360}{504}$	0.029	$-\frac{504}{720}$	0.022	$-\frac{618}{990}$	0.030
		$-\frac{96}{120}$	0.067	$-\frac{174}{210}$	0.029	$-\frac{200}{336}$	0.017	$-\frac{396}{504}$	0.014	$-\frac{552}{720}$	0.011	$-\frac{678}{990}$	0.017
		$-\frac{108}{120}$	0.042	$-\frac{186}{210}$	0.017	$-\frac{212}{336}$	0.0062	$-\frac{432}{504}$	0.0054	$-\frac{600}{720}$	0.0041	$-\frac{738}{990}$	0.0087
		-1	0.0083	$-\frac{198}{210}$	0.0083	$-\frac{224}{336}$	0.0014	$-\frac{468}{504}$	0.0011	$-\frac{648}{720}$	0.0010	$-\frac{798}{990}$	0.0036